Submitted to *Operations Research*
manuscript (Please, provide the mansucript number!)

Modeling and Design of Container Terminal Operations

Debjit Roy
Indian Institute of Management, Ahmedabad, India, debjit@iimahd.ernet.in

René de Koster
Rotterdam School of Management, Erasmus University, The Netherlands, RKoster@rsm.nl

Design of container terminal operations is complex because multiple factors affect the operational performance. These factors include: topological constraints, a large number of design parameters and settings, and stochastic interactions that interplay among the quayside, vehicle transport, and stackside processes. In this research, we propose new integrated queuing network models for rapid design evaluation of container terminals with Automated Lift Vehicles (ALVs) and Automated Guided Vehicles (AGVs). These models offer the flexibility to analyze alternate design variations and develop insights. For instance, the effect of alternate vehicle dwell point policy is analyzed using state-dependent queues, whereas the efficient terminal layout is determined using variation in the service time expressions at the stations. Further, using embedded Markov chain analysis, we develop an approximate procedure for analyzing bulk container arrivals. These models form the building block for design and analysis of large-scale terminal operations. We test the model efficacy using detailed in-house simulation experiments and real-terminal validation by partnering with an external party.

Key words: Container terminal, intra-terminal transport, design decisions, queuing models

1. Introduction

Due to growth in international trade and better accessibility to the major seaports via deep-sea vessels, containerization has become a preferred mode for maritime shipping and inland transportation. With over 90% of the global trade carried over sea, the maritime containerization market is projected to reach 731 Million TEU by 2017 (Global Industry Analysts Inc. (2013)). Currently, several new deep-sea as well as hinterland automated container terminals are being designed across continents.

During the terminal design process, the terminal operator makes strategic design decisions such

Electronic copy available at: https://ssrn.com/abstract=2458048

as deciding the berthing capacity, layout of the terminal, type of equipment (quay crane, stacking crane) for handling containers at the seaside and landside, and type of vehicle (Automated Lift Vehicle, ALV with self-lifting capability vs. Automated Guided Vehicle, AGV without self-lifting capability, see Figure 1) for container transport between seaside and the landside. They also address tactical issues such as berth allocation, container stowage, vehicle routing as well as operational design choices such as vehicle dispatching policy and dwell point policies (see Günther and Kim (2006), Steenken et al. (2004)).

(a) (b)

Figure 1 a) Automated Lift Vehicle (ALV) (source: *www.sae.org*), b) Automated Guided Vehicle (AGV)

Making the right design decision is crucial as the investments involved are huge (between 1 and 5 billion euros) and the time to realize is large (in many cases the land and the port has to be created), and the payback period varies between 15-30 years (Wiegmans et al. (2002)). The process to arrive at a high-performing container terminal design configuration is complex due to multiple factors: 1) *physical constraints*: variations in ground conditions and topology of the terminal area, 2) *large design search space*: large number of design parameters that affects operational performance, and 3) *stochastic interactions*: the ship arrival, container handling, and vehicle transport processes are integrated with each other. For example, in the container unload operation, the departure process outputs from the quay cranes form the arrival process inputs to the vehicle transport. Since design

Electronic copy available at: https://ssrn.com/abstract=2458048

of a container terminal has a significant effect on the operational performance, analyzing the design trade-offs and identifying an efficient operating range for the design parameters using stochastic models are of immense interest to the terminal operators.

Existing research on modeling and design of container terminal operations can be broadly classified into two categories depending on the type of system under investigation: 1) isolated system, where the system under consideration is one among the three processes: quayside, vehicle transport, or stackside operations, and 2) integrated system, where the three processes and their interactions are studied together as one integrated operation. We first discuss the related work on isolated systems.

One of the research focuses on isolated systems has been on developing optimization and simulation models to address *operational issues* such as scheduling of container storage and retrieval operations (Vis and Roodbergen (2009), Gharehgozli et al. (2012)), real-time yard truck and yard crane control (Petering et al. (2009), Petering and Murty (2009), Petering (2010)), routing algorithms for transfer cranes (Kim and Kim (1999)), quay crane scheduling (Kim and Park (2004), Liang et al. (2009)), and workload management at the yard cranes (Ng (2005), Petering (2011)). The second stream of research focuses on evaluating *design decisions* of isolated systems. Using detailed simulation models, researchers have studied the performance and cost trade-offs using different type of vehicles for inter-terminal container transport: multi-trailers, automated guided vehicles or automated lift vehicles (see Duinkerken et al. (2007), Vis and Harika (2004)). Performance analysis of specific container terminal design aspects has also been carried out using stochastic models, which are evaluated rapidly. For an overview of literature on container terminal modeling, refer Vis and De Koster (2003) and Steenken et al. (2004). We discuss the literature based on two broad functional areas: quayside, and vehicle transport and stackside.

Quayside operations: Koenigsberg and Lam (1976) developed a closed queuing network model for a multi-port system and estimate performance measures such as the expected number of vessels waiting in each stage, and, most important, the expected waiting time in port. Easa (1987) presented approximate queuing models to help assess the impacts of tug services on congested

Electronic copy available at: https://ssrn.com/abstract=2458048

harbor terminals. A congested harbor terminal is modeled as a queuing system with m identical tugs as servers and n identical berths as customers, and with general probability distributions of tug service time and berth cargo-handling time. Dragovic et al. (2006) developed performance evaluation of ship-berth link using an $M/E_k/n_b$ model where n_b is the number of berths at the port. Further, they experiment the model with data based of Pusan East Container Terminal (PECT), Korea. Mennis et al. (2008) developed a continuous time Markov chain (CTMC) model to analyze the loading and unloading procedure of a ship with the use of gantry cranes. The state description of the model captures the operating state of the crane such as available, failure, waiting for repair, and replacement. Canonaco et al. (2008) developed a queuing network model to analyze the container discharge and loading at any given berthing point. However, they evaluate the network using simulation.

Vehicle transport and stackside operations: Guan and Liu (2009) developed a multi-server queuing model to analyze gate congestion for inbound trucks. Further, they determined the optimal number of gates to minimize the sum of truck waiting costs and the gate operating costs. Li et al. (2009) developed a discrete time model to schedule two stacking cranes that process the storage and retrieval requests in a single block with an I/O point located at one side of the block along the bays. The cranes cannot pass each other and must be separated by a safety distance. Vis and Carlo (2010) also considered a similar problem; however, in their problem the cranes can pass each other but cannot work on the same bay simultaneously. They formulated the problem as a continuous time model and minimize the makespan of the stacking cranes. Gharehgozli et al. (2012) introduced a continuous time model to schedule two interacting stacking cranes that work in a single block of containers. Their model incorporates several types of constraints such as precedence constraints, crane interaction constraints, and constraints that assign each container to a storage location selected from a given set.

We now discuss existing work on integrated systems, which is also the focus of our research. Simulation models are developed to analyze operational rules such as the effect of vehicle dispatching policies (De Koster et al. (2004)). They developed the model for a terminal in the Port of

Electronic copy available at: https://ssrn.com/abstract=2458048

Rotterdam and analyze alternate dispatching rules such as nearest vehicle first and nearest vehicle with time priority on throughput times. Integrated optimization models are also developed to solve berth allocation, quay crane assignment, and quay crane scheduling problems in a unified manner. Meisel and Bierwirth (2013) provides a framework for aligning the three decisions in an integrated fashion whereas Vacca et al. (2013) present an exact branch and price algorithm for solving the berth allocation problem along with the quay crane assignment.

Analytical and simulation models are also developed to analyze terminal design decisions. For instance, Hoshino et al. (2004) proposed an optimal design methodology for an Automated Guided Vehicles (AGV) transportation system by using a combination of closed queuing network and a simulation model. Bae et al. (2011) compared the operational performance of an integrated system with two types of vehicles (ALVs and AGVs). Through simulation experiments they show that the ALVs reach the same productivity level as the AGVs using much less number of vehicles due to its self-lifting capability. Practitioners primarily develop detailed simulation models to design new terminals or improve the efficiency of existing terminal operations (see TBA BV (2010)). While simulations can help for detailed analysis, the complexity and interactions involved make such models prohibitively expensive and time consuming if used for generating and selecting designs (Edmond and Maggs (1976)).

The review of related work highlights several gaps. First, existing research on modeling and design of stackside and vehicle transport operations using analytical models is limited. Deciding the optimal terminal layout, for instance, the optimal number of rows/stack and number of bays are important design decisions. Further, this decision is tightly coupled with the vehicle transport because the number of stack blocks affects the length of the vehicle travel path. In the vehicle transport area, the type of vehicle and the dwell point (parking location of the vehicles) affect the throughput performance. Note that a terminal with AGVs is a tightly coupled system, where quay crane and AGVs at the quayside, and AGVs and stack crane at the stackside, need to be synchronized for the transfer of a container at the quayside and stackside respectively. Second, existing models analyze specific terminal subprocess such as ship-berth interface, which limits

Electronic copy available at: https://ssrn.com/abstract=2458048

its practical applicability in making integrated terminal design decisions. Further, bulk arrival of containers affects the waiting time downstream due to stochastic interactions among the quayside, transport, and stackside processes, emphasizing the need to analyze integrated operations. Finally, for modeling simplicity, researchers have assumed the topology of the vehicle travel path to be uni-directional loop without any shortcuts. Shortcuts, which are used in practice, can significantly improve vehicle flow times and minimize congestions.

To bridge the gaps in research, we develop new integrated queuing network models for the unloading and loading of containers at the seaside by considering the stochastic interactions among the quayside operations, vehicle transport operations, and stackside operations. First, we develop an elaborate model with ALVs to capture the congestion effects at the quayside, transport, and stackside processes. Each quay crane is modeled as a $GI/G/1$ queue. For container transport in the yard area, the containers may wait for an available vehicle. However, due to the capacity constraints of the quay crane, an ALV may also wait for a container arrival. This interaction between ALVs and containers is precisely modeled using a synchronization station and the queuing dynamics in the vehicle transport is modeled using a semi-open queuing network (SOQN) with V ALVs. Each stack crane is also modeled as a $GI/G/1$ queue. The individual models are linked using a parametric decomposition approach. Next, we develop the integrated queuing network model for the container unloading and loading processes using AGVs. To model the hard coupling of AGVs with quay and stack crane resources, we develop synchronization protocols for the quayside and the stackside processes. Use of the two protocols reduces the model complexity and allows to capture the congestion effects in the AGV-based system as a single SOQN. Since the SOQN is not product form, we develop an approximate solution procedure for network evaluation. Using the two integrated models and their extensions, we answer the following research questions:

1. Which one is the *better horizontal transport system,* a decoupled system with ALVs or a coupled system with AGVs? While AGVs require less container loading/unloading times (due to synchronization), the required number of AGVs may be more to achieve a target throughput capacity in comparison to ALVs.

Electronic copy available at: https://ssrn.com/abstract=2458048

2. What is the *optimal terminal layout configuration* combination of the number of stacks, number of lanes/stack, and number of tiers/stack, which minimizes the expected throughput times?

3. What is the effect of *vehicle dwell point* on expected throughput times? A vehicle may dwell at a parking location, if no containers are waiting to be processed. From known warehousing literature, the resource dwell point has less impact at high utilization (Meller and Mungwattana (2005)). Does the results hold true for container terminals too?

4. What is the effect of *bulk arrival of containers* on expected throughput times?

Our work closely aligns with the analytical model developed by Hoshino et al. (2004). However, our research contributes to the stochastic modeling and transportation system modeling and design literature in several aspects: 1) we develop a semi-open queuing network model of the terminal system, which considers the synchronization of the AGVs and the containers waiting at the vessel to be unloaded. We use a semi-open queuing network to realistically capture the effect that sometimes an AGV will be waiting for a container to be unloaded while during other times, a container will be waiting in the vessel for unloading operations, 2) we develop protocols for handling containers at the quayside and the stackside that allows us to models the vehicle synchronization effects at the quay and the stack area, 3) we consider a vehicle travel path topology with multiple shortcuts that decreases the average travel times and improves vehicle capacity. Previous models do not consider the effects of multiple shortcuts, 4) we develop a state-dependent semi-open queuing network model to analyze the effect of alternate dwell point policies, 5) we adapt our model to analyze alternative terminal layouts by varying the number of stacks, bays, and vehicle path dimensions, and arrive at a layout that minimizes throughput times. Our model is flexible because it does not impose any restriction on the type of transport layout. The distribution of the travel times in the vehicle travel can be altered to mimic the type of transport layout. Using analytical approximations, we analyze the tradeoffs between the number of ALVs and AGVs required to achieve a throughput, and tradeoffs between the vehicle transport and the SC movement times. Further, with a model variation, we develop design insights with respect to vehicle dwell point policy.

Electronic copy available at: https://ssrn.com/abstract=2458048

The rest of this paper is organized as follows. The container terminal layout and the system configuration is described in Section 2 whereas the model assumptions and the queuing network model for the unloading operations with ALVs is presented in Section 3. The analytical model for the unloading operations with AGVs is presented in Section 4. Using variations of the analytical model with ALVs, we develop insights with respect to the vehicle dwell point policy and bulk container arrivals. These models are described in Section 5. The models are validated using detailed simulation models. The numerical experiments and the design insights with respect to performance comparison between ALV and AGV-based systems, effect of vehicle dwell point, and optimal terminal layout are presented in Section 6. Finally, we summarize our key findings and model extensions in Section 7.

2. Terminal Layout and Seaside Operations

This section describes the container handling (loading and unloading) processes at the seaside, illustrates the topology of the vehicle travel path, and finally presents the integrated terminal layout considered in this research.

2.1. Scope and Container Handling Process

The terminal area is composed of two sections: seaside and landside. The seaside area includes the quay area, transport area, and the stack area, which are operated by a fleet of quay cranes (QCs), vehicles and the stack cranes (SCs) respectively (see Figure 2, Brinkmann (2010)). For the terminal operator, seaside operations are critical because the shipping liners, who are the paying customers, select terminals that offer quickest service at the lowest cost. Costs of other activities should be indirectly covered from the fees of the shipping liners. This research, therefore, focuses on the seaside operations, which entails unloading of containers from the import vessel and loading of containers to the export vessel.

The process of loading and unloading containers on the vessel is dependent on the type of vehicle used for horizontal transport. Since the ALVs have the capability to self-lift containers, they do not need to be synchronized with the QC or the SC for loading/unloading of containers. In contrast,

Electronic copy available at: https://ssrn.com/abstract=2458048

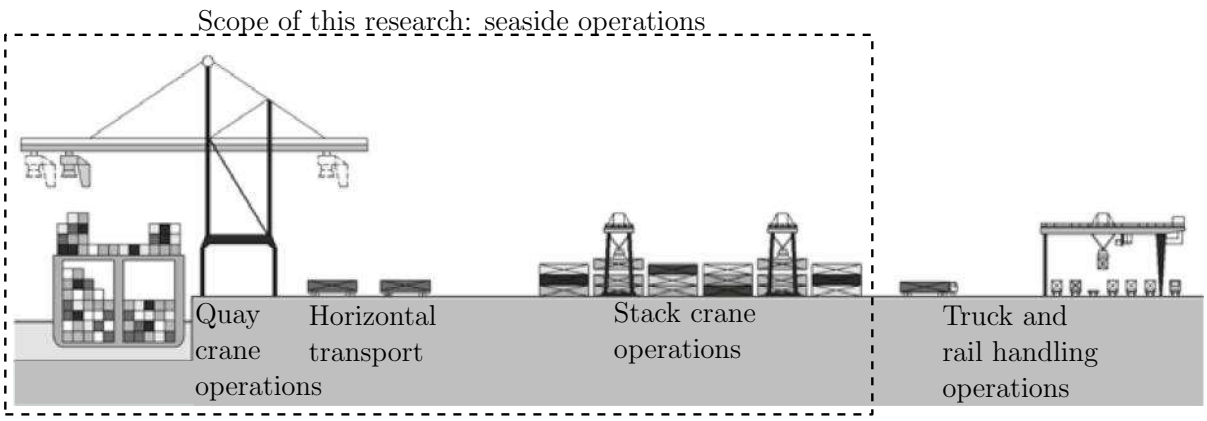

Figure 2 General layout of a container terminal and scope of this research (Adapted from Brinkmann (2010))

AGVs do not have the capability to self-lift containers and hence, the QC and SC operations need to be synchronized with the AGVs. We first develop the queuing network models by considering ALVs for horizontal transport, which decouple the loading and unloading of containers from the QCs and the SCs. The vehicle transport using AGVs is considered later. Further, we provide a detailed description of the queuing models with respect to the unload operations.

The container unload operation with ALVs is composed of three steps including quayside, vehicle transport, and stackside processes. In the quayside process, the QCs unload the containers from the ship and position them at a buffer area. From the buffer area, the ALVs transport the containers to the stack buffer location (vehicle transport process). The SCs move the containers from the stack buffer location and store them at a stack location (stackside process). The process steps involved in the loading of containers are similar to the unloading operations except that the steps are executed in a reverse order (see Figure 3). The throughput time expressions (CT_u and CT_l) to unload and load a container using ALVs include both the waiting times as well as the movement time for the three processes (Equations 1 and 2).

$$CT_u = W_q + T_q + W_v + T_v + W_s + T_s \tag{1}$$

$$CT_l = W_s + T_s + W_v + T_v + W_q + T_q \tag{2}$$

The components W_q, W_v, and W_s denote the waiting times for the QC, vehicle, and SC respectively whereas the components T_q, T_v, and T_s denote the container movement time by the QC,

Electronic copy available at: https://ssrn.com/abstract=2458048

vehicle, and SC respectively. Modeling a terminal system that accounts for the service time variabilities in QC operations, internal transport operations (due to variation in the origin and the destination location), and SC operations, along with stochastic interactions among the processes and disturbances in the system such as manual intervention in QC operation, delays associated with manual twist lock removal in the vessel, may be intractable using a deterministic model with delays. Hence, an integrated queuing network model is useful to capture the interactions among the processes and develop design insights.

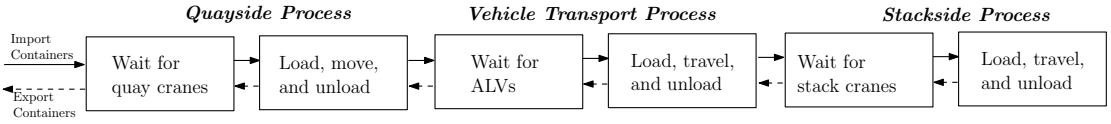

Figure 3 Illustration of container loading and unloading process with ALVs (excludes ship berthing and landside operations)

2.2. Topology of Vehicle Transport Path

A significant time is spent in transporting containers from the stackside to the quayside and vice versa. The travel time of a vehicle depends on multiple parameters such as the originating point of the vehicle, destination point of the vehicle, dwell point location of the vehicle, and the layout of the transport path. Among these parameters, the layout of the transport path is an important design choice because it not only demands substantial investments and infrastructure but is also influenced by the number of the stacks, rows/stack, and the QC locations.

Figure 4 illustrates three layouts of automated guided vehicles used in the container terminal research. Figure 1(a) illustrates a single unidirectional travel layout used by all vehicles (see Vis and Harika (2004)). However, multiple parallel travel paths at the quayside (corresponding to the different buffer locations) as well as stackside may be preferred to minimize congestion and permit differential path travel velocity (high speed path vs low speed path) (see Hoshino et al. (2004)). In addition, multiple shortcuts are present at the terminals to minimize the travel time between the quayside and the stackside area (see De Koster et al. (2004)). In this research, we adopt layout

Electronic copy available at: https://ssrn.com/abstract=2458048

3, which is used at various terminals in Europe. This layout includes shortcut paths, which can have a substantial effect on minimizing travel times (especially from quayside to the stackside) and improving system performance. However, our model permits the analysis of other design layouts for vehicle transport.

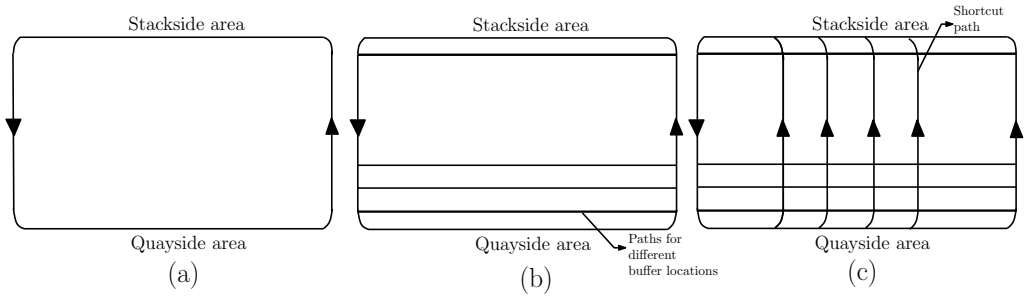

Figure 4 Types of guide-path used in the transport area: a) single unidirectional loops, b) multiple unidirectional loops, and c) multiple unidirectional loops with shortcuts between quayside and stackside area

2.3. Integrated Terminal Layout

Figure 5 depicts the top view of a part of a container terminal, which includes the quayside, transport and the stackside area (QCs, transport area with vehicles, stack blocks with cranes). The design of this layout is motivated from practice (see De Koster et al. (2004)). We focus on the space allowing berthing of one jumbo vessel with a drop size of several thousands of containers. A large container terminal may contain several of such identical berthing positions. The number of stacks is denoted by N_s and each stack crane is referred as SC_i where $i : i \in \{1, \ldots, N_s\}$. Similarly, the number of QCs is denoted by N_q and each crane is referred as QC_j where $j : j \in \{1, \ldots, N_q\}$. There is one shortcut path after each QC (referred as SP_j where $j : j \in \{1, \ldots, N_q\}$) that connects the quayside to the stackside areas. Both stacks and QCs have a set of buffer lanes, which are used by the vehicles to park during loading or unloading containers. The number of buffer locations at each QC and SC are denoted by N_{qb} and N_{sb} respectively. The other notations present in Figure 5 indicate path dimensions, which are used later to estimate the vehicle travel times. The next section presents the integrated queuing network model for the unloading and loading operations using ALVs.

Electronic copy available at: https://ssrn.com/abstract=2458048

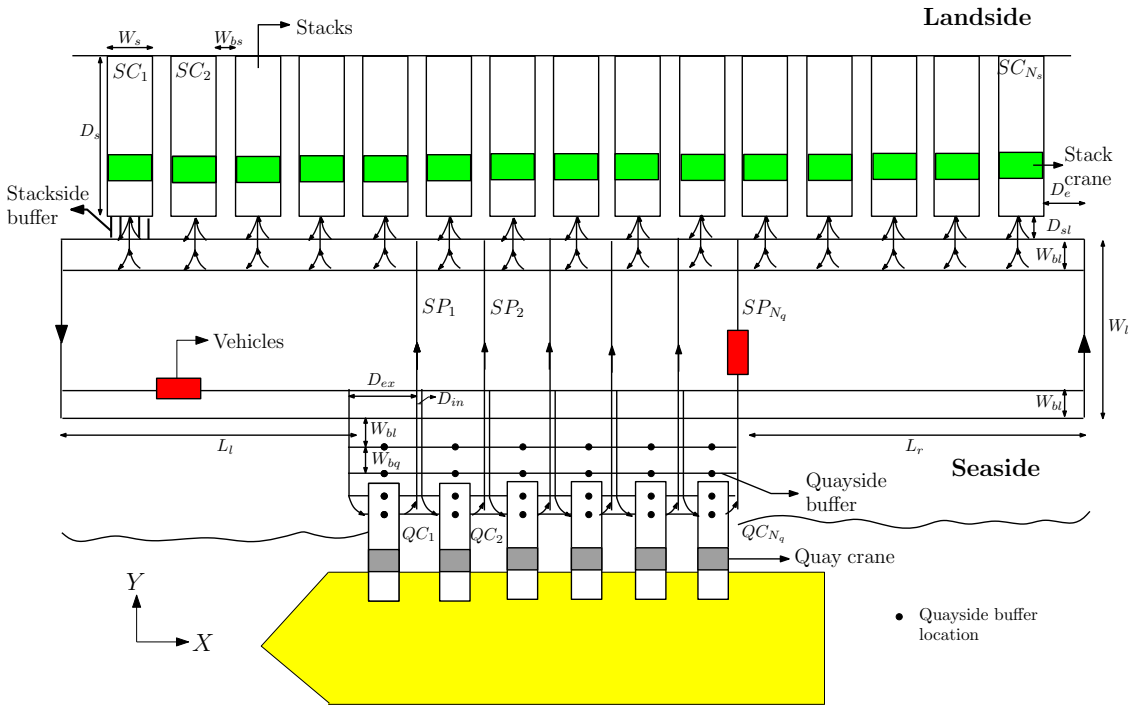

Figure 5 Layout of the container terminal used in this research

3. Queuing Network Model for Terminal Operations with ALVs

In this section, we first describe the model assumptions and then develop the models for the three processes: quayside, vehicle transport, and stackside. The last section describes the integrated network model, which is obtained by linking the arrival and departure processes of the three process models using a parametric decomposition approach.

3.1. Model Assumptions

As described earlier, the container loading/unloading process is composed of three sub-processes: quayside, vehicle transport, and stackside. Typically, a set of QCs, vehicles, and SCs are dedicated to either loading or unloading containers. For instance, when an import ship arrives, the QCs are dedicated for the unloading operations. When a portion of the ship is unloaded, then the QC is dedicated for the loading operations. Hence, we analyze the two processes (loading and unloading) separately. The modeling is first described with respect to the unloading operations. We now list the modeling assumptions for the three processes.

Electronic copy available at: https://ssrn.com/abstract=2458048

Quayside process: Some modern QCs have two trolleys that work in tandem. For modeling simplicity, we assume that there is one trolley per QC. There is infinite buffer space for parking vehicles at the QC location. Note that in the real situation, the buffer space is finite. We assume that vehicles can park at locations near the QC buffer space. Hence this assumption does not limit the utility of the model. The dwell point of cranes is the point of service completion. We assume a single container arrival flow (relaxed later to bulk arrivals, see Appendix E) and random assignment of containers to QC. Note that when a vessel arrives, all containers are present but not all are available for pickup. The containers can be unloaded only in a sequence, i.e., the containers on the top need to be unloaded before another container becomes available. Also, deck hatches may have to be removed and containers need to be manually unlocked before the QC can pick them. This activity adds variability in the inter-arrival times of the containers. This randomness can very well be modelled using a stochastic arrival distribution. We assume a Poisson arrival process; however, our model permits other container arrival and assignment distributions.

Vehicle transport process: The type of vehicles is ALV (relaxed later to AGV). Each vehicle can transport only one container at a time. The dwell point of the vehicles is the point of service completion i.e., the vehicle dwells at the stackside buffer lane after completing the unload operation (relaxed later to other parking locations). The vehicle dispatching policy is FCFS. The vehicle travel paths are unidirectional with multiple shortcuts (layout 3). Blocking among vehicles at path intersections is not considered and the model assumes a uniform vehicle velocity. Further, vehicle acceleration and deceleration effects are ignored.

Stackside process: The number of storage locations is fixed; we vary the number of stacks (N_s), number of rows per stack (N_r), bays per stack (N_b), and tiers per stack (N_t). Further, the storage location of a container, which is uniquely defined by a combination of four variables: stack number, row number, bay number and tier number, is selected randomly. In automated stacking

Electronic copy available at: https://ssrn.com/abstract=2458048

crane terminals (using Rail Mounted Gantry Cranes), the workload is typically distributed over many stacking cranes to serve the QCs in parallel (Saanen and Dekker (2007)). Further, random stacking strategy is also adopted for benchmarking the performance of different stacking strategies (Dekker et al. (2006)). The dwell point of cranes is the point of service completion. Similar to the quayside, we also assume infinite buffer space for parking vehicles at the SC location.

While our assumptions are abundant, most of them can be relaxed, albeit at the expense of more complicated modeling.

3.2. Model Description

From Figure 3, it can be seen that for the unload operations, the three processes quayside, transport, and stackside processes are linked with each other. The container departure information from the quayside provides the arrival process inputs to the vehicle transport. Similarly, the container departure information from the vehicle transport forms the arrival process inputs to the stackside. Hence, we develop three models corresponding to the three processes and determine the performance measures and departure process information for each of them. Using the arrival and departure process information for the individual queuing models, we develop the integrated model of the container unload operation.

3.2.1. Quayside Process. At the quayside, the containers that wait to be unloaded, queue at one of the QCs. Once a crane is available, the total time taken by the QC to unload a container from the vessel to a QC buffer location includes pick-up, move, and dropoff time. The queuing analysis is discussed now.

The objective of the QC queuing model is to estimate the performance measures, and the squared coefficient of variation (SCV) of the inter-departure times from the QC resources ($c^2_{d_{q_i}}$). The inputs to the model are the first moment and the SCV of the inter-arrival times to the QCs, and the QC service times. Let the time to unload a container from the ship using a QC i be a random variable, $\mathcal{X}_{q_i}$, with mean $\mu^{-1}_{q_i}$ and squared coefficient of variation, $c^2_{s_{q_i}}$ where $i = \{1, \ldots, N_q\}$. Further,

Electronic copy available at: https://ssrn.com/abstract=2458048

the mean and the squared coefficient of variation for the container inter-arrival times to the QCs are denoted by parameters $\lambda_{a_{q_i}}^{-1}$ and squared coefficient of variation, $c_{a_{q_i}}^2$. Let λ_a be the overall container arrival rate; due to the thinning process, the arrival process to each QC is also Poisson with rate $\frac{\lambda_a}{N_q}$. With these input parameters, each QC queue is modeled as a $GI/G/1$ queue and the performance measures such as expected throughput time ($\mathbb{E}[T_q]$), crane utilization ($\mathbb{U}_{q_i}$) and the SCV of the inter-departure times are estimated using two-moment approximation results, Whitt (1983) (Equations 3-4).

$$\mathbb{W}_q(i) = \left(\frac{\mu_{q_i}^{-1}\mathbb{U}_{q_i}}{1-\mathbb{U}_{q_i}}\right)\left(\frac{c_{a_{q_i}}^2 + c_{s_{q_i}}^2}{2}\right) \tag{3}$$

$$c_{d_{q_i}}^2 = \mathbb{U}_{q_i}^2 c_{s_{q_i}}^2 + (1-\mathbb{U}_{q_i}^2)c_{a_{q_i}}^2 \tag{4}$$

where $i = 1, \ldots, N_q$ and $\mathbb{U}_{q_i} = \frac{\lambda_{a_{q_i}}}{\mu_{q_i}}$

3.2.2. Vehicle Transport Process. The number of ALVs in the system is denoted by V. ALVs transport containers between quayside and the stackside, using defined guide paths. The outer guide path (see Figure 5) is used by vehicles when they approach the stack or the quayside buffer area, while the inner lane is used by vehicles during intermediate travel. Note that by using two paths, the vehicle congestion within a single path is reduced. The objective of the vehicle transport queuing model is to estimate the performance measures, and SCV of the inter-departure times from the ALV network ($c_{d_t}^2$). The input parameters to the ALV queuing model are the mean and the SCV of the container inter-arrival times ($\lambda_{a_t}^{-1}$ and $c_{a_t}^2$), and the mean and SCV of the vehicle service times (μ_t^{-1} and $c_{s_t}^2$). We first describe the approach to estimate the travel time distribution and then present the queuing model for vehicle transport.

During the unloading process, vehicles dwell at the point of service completion (stackside buffer locations) after unloading containers. Let the random variable, $\mathcal{X}_t$, denote the service time to complete one travel cycle (see Equation 5).

$$\mathcal{X}_t = \mathcal{X}_{sq} + \mathcal{X}_{lu} + \mathcal{X}_{qs} \tag{5}$$

Electronic copy available at: https://ssrn.com/abstract=2458048

where $\mathcal{X}_{sq}$, $\mathcal{X}_{lu}$, and $\mathcal{X}_{qs}$ denote the random variables corresponding to the travel between stackside and quayside, load/unload times, and travel time between quayside and stackside respectively. We now describe the procedure to estimate the mean and variance of $\mathcal{X}_t$.

Let μ_t^{-1} denote the mean service time to complete one travel cycle, i.e, the cumulative sum of the expected travel time from stackside to the quayside (τ_{qs}^{-1}), deterministic container pickup and drop time (L_t^v and U_t^v), and expected travel time from quayside to the stackside (τ_{sq}^{-1}). Note that we consider shortest path route information (from origin to destination location) to develop the service time expressions . Therefore, μ_t^{-1}, includes the minimum expected travel time required to travel from origin (quayside to stackside and return). The notations used in the service time expression for the vehicle transport are described in Table 1.

Table 1 Notations used in the service time expressions for the vehicle transport (Refer Figure 5)

Term	**Description**
N_q	Number of QCs
N_b	Number of buffer locations per QC
N_s	Number of stacks
W_s	Width of a stack
W_{bs}	Width between stacks
D_e	Distance between last stack along the X-axis (both ends)
W_{bl}	Distance between two tracks
W_{bq}	Distance between two buffer lanes at the quayside
D_{ex}	Distance between entrance and exit of each shortcut
D_{in}	Distance between exit of one shortcut and entrance of another shortcut
D_{sl}	Length of the buffer lane at the stackside
L_r	Length of the path after the last shortcut
L_l	Length of the path before the first shortcut
W_l	Width of the vehicle path
$N_{sq}(i)$	Number of shortcut paths corresponding to QC i
$N_{sl}(i)$	Number of stack blocks after the shortcut of each QC i (to the right)
L_t^v, U_t^v	Container loading and unloading time
h_v	Vehicle velocity

The travel time from the QC to a stack depends on the relative location of the stack with respect to the QC. For instance, if the stack is located towards the far right and does not have a direct access to a shortcut path, then the vehicle travels longer to unload a container. Based on the location of the QC and the stack, three cases are developed to estimate the first and second

Electronic copy available at: https://ssrn.com/abstract=2458048

moment of the travel time from QC to the stack buffer lane. Let index $i \in \{1, \ldots, N_q\}$ refer a QC whereas index $j \in \{1, \ldots, N_s\}$ refer an SC. In case 1, the destination stack (SC_j) is located to the left of the shortcut path ($SP_i : i \in \{1, \ldots, N_q\}$) whereas the destination stack is located between $SP_i : i \in \{1, \ldots, N_q - 1\}$ and SP_{N_q} in case 2. Note that case 2 does not apply for the last QC because the stacks are located at either left or right of this QC. In case 3, the destination stack is located after the last shortcut path, SP_{N_q}. See Equations 27-29 for the casewise expressions and Equation 30 for the final travel time expression from quayside to stackside (Appendix A). Note that we assume that each shortcut path ends at the midpoint between the end of one stack and the beginning of another stack. Further, the number of stacks enclosed between two consecutive shortcuts is one. Also, the travel time expressions are developed with respect to the stack buffer lane located at the middle of its stack. In case 1, the vehicle travels from the QC i along the quayside $\left(\frac{D_{ex}}{2h_v} + \frac{W_{bl}+((N_b-1)W_{bq})/2}{h_v}\right)$ time units, then travels along the shortcut path to the stackside $\left(\frac{W_l}{h_v}\right)$ time units, and then travels along the stackside to reach the destination stack j's buffer lane $\left(\frac{W_{bs}}{2h_v} + (N_s - N_{sl}(i) - j)\left(\frac{W_s+W_{bs}}{h_v}\right) + \frac{W_s}{2h_v} + \frac{D_{sl}}{h_v}\right)$ time units. Similarly, the expected travel time from the stack buffer lane to the QC buffer lane(τ_{sq}^{-1}) is estimated using Equation 31. Note that, since the shortcut paths are unidirectional, there exists only one unidirectional path from a stack to a QC. The service time expressions are included in Appendix A.

From Equation 5, the final expression of the expected vehicle travel time, μ_t^{-1}, is given by Equation 6.

$$\mu_t^{-1} = \tau_{sq}^{-1} + \tau_{lu}^{-1} + \tau_{qs}^{-1} \tag{6}$$

where $\tau_{lu}^{-1} = L_t^v + U_t^v$. Since, the random variables are assumed independent of each other, the second moment of the service time is determined using the relation $\mathbb{E}[\mathcal{X}_{sq} + \mathcal{X}_{lu} + \mathcal{X}_{qs}]^2$, and the SCV of the service time ($c_{s_t}^2$) is estimated using the relation $\frac{\mathbb{E}[x_{sq}+x_{lu}+x_{qs}]^2 - (\mathbb{E}[x_{sq}+x_{lu}+x_{qs}])^2}{(\mathbb{E}[x_{sq}+x_{lu}+x_{qs}])^2}$

With the known input parameters, the vehicle transport is modeled as follows. For container transport in the yard area, the containers may wait for an ALV. However, due to the capacity constraints of the QC, an ALV may also wait for a container arrival. This interaction between ALVs

Electronic copy available at: https://ssrn.com/abstract=2458048

and containers is precisely modeled using a synchronization station and the queuing dynamics in the vehicle transport is modeled using a semi-open queuing network (SOQN) with V vehicles circulating in the network. The containers, which are unloaded from the vessel by the QCs, queue at the quayside buffer locations (designated as buffer B_1 in the queuing model) to be picked up by the ALVs. Idle vehicles queue at buffer location (B_2) to process a transaction. Once an ALV and a container is available, the ALV queues at an Infinite Server (IS) station to receive service. After service completion, the ALV queues at stackside buffer locations (designated as buffer B_2 in the queuing model). The approach to estimate the network performance measures, and the SCV of the inter-departure times from the IS station is explained in Section 3.2.4.

3.2.3. Stackside Process. As discussed earlier, N_s, N_b, N_r, and N_{bs} denote the number of SCs, number of bays/stack, number of rows/stack, and number of buffer lanes/stack. At the stackside, the containers that wait to be stored, queue at one of the SCs' buffer lanes. Once a crane is available, the total time taken by the SC to store a container includes move time from crane dwell point to pick-up location, container pickup time, and move and dropoff time. The queuing analysis is discussed now.

The objective of the SC queuing model is to estimate the performance measures. The inputs to the model are the first moment and the SCV of the container inter-arrival times to the SC queue ($\lambda_{a_{s_i}}^{-1}$ and $c_{a_{s_i}}^2$), and the mean and SCV of the crane service times. The mean inter-arrival time to each crane, $\lambda_{a_{s_i}}^{-1}$, is $\left(\frac{\lambda_a}{N_s}\right)^{-1}$.

During the unloading process, stack cranes dwell at the point of service completion (stack storage locations) after unloading containers. Let the random variable, $\mathcal{Y}_s$, denote the service time to complete one travel cycle (see Equation 7).

$$\mathcal{Y}_s = \mathcal{Y}_{sb} + \mathcal{Y}_{lu} + \mathcal{Y}_{bs} \tag{7}$$

where $\mathcal{Y}_{sb}$, $\mathcal{Y}_{lu}$, and $\mathcal{Y}_{bs}$ denote the random variables corresponding to the horizontal travel time between stack dwell point location to the container pickup location at the stack buffer lane, vertical

Electronic copy available at: https://ssrn.com/abstract=2458048

container pick-up and container dropoff time, and horizontal travel time from stack buffer lane pickup location to the container dropoff point in the storage area. We now describe the procedure to estimate the mean and variance of $\mathcal{Y}_s$.

Due to the random location selection for container storage and the point of service completion dwell point of the SC, the originating and destination location of an SC in each cycle follows uniform distribution. Also, the container pickup location (stack buffer lanes) is selected with equal probability. We account for the vertical travel time of the crane in the pickup and dropoff time. Let x_{n_i}, y_{m_i} and x_{n_j}, y_{m_j} denote the origin and the destination coordinates of the crane. Due to simultaneous movement of the crane in the x and y direction, $max\left(\frac{|x_{n_i}-x_{n_j}|}{v_{s_x}}, \frac{|y_{m_i}-y_{m_j}|}{v_{s_y}}\right)$ denotes the effective horizontal travel time, where v_{s_x} and v_{s_y} denote the crane velocity along the x- and the y-axis respectively (see Figure 6). Let L_t^s and U_t^s denote the container pickup and drop times.

The service times also have a general distribution with mean $\mu_{s_i}^{-1}$, which is dependent on the travel trajectory of the crane. Since, the random variables are assumed independent of each other, the second moment of the service time is determined using the relation $\mathbb{E}[\mathcal{Y}_{sb} + \mathcal{Y}_{lu} + \mathcal{Y}_{bs}]^2$ and the SCV of the service time ($c_{s_s}^2$) is estimated using the relation $\frac{\mathbb{E}[\mathcal{Y}_{sb}+\mathcal{Y}_{lu}+\mathcal{Y}_{bs}]^2-(2\mathbb{E}[\mathcal{Y}_{sb}]+\mathbb{E}[\mathcal{Y}_{lu}])^2}{(2\mathbb{E}[\mathcal{Y}_{sb}]+\mathbb{E}[\mathcal{Y}_{lu}])^2}$. With these inputs, each SC resource is also modeled as a $GI/G/1$ queue where the inter-arrival times are independent and identically distributed. The SC throughput time is denoted as $\mathbb{E}[T_s]$.

$$\mathbb{E}[\mathcal{Y}_{sb}] = \sum_{n_i=1,m_i=1}^{n_i=N_r,m_i=N_b} \sum_{n_j=1,m_j=1}^{n_j=N_{bs},m_j=1} \frac{1}{(N_b N_r N_{bs})} max\left(\frac{|x_{n_i}-x_{n_j}|}{v_{s_x}}, \frac{|y_{m_i}-y_{m_j}|}{v_{s_y}}\right) \tag{8}$$

$$\mathbb{E}[\mathcal{Y}_{lu}] = L_t^s + U_t^s \tag{9}$$

Note that $\mathbb{E}[\mathcal{Y}_{sb}] = \mathbb{E}[\mathcal{Y}_{bs}]$, therefore $\mu_{s_i}^{-1}$ can be expressed as follows:

$$\mu_{s_i}^{-1} = 2\mathbb{E}[\mathcal{Y}_{sb}] + \mathbb{E}[\mathcal{Y}_{lu}] \tag{10}$$

$$\mathbb{W}_s(i) = \left(\frac{\mu_{s_i}^{-1}\mathbb{U}_{s_i}}{1-\mathbb{U}_{s_i}}\right)\left(\frac{c_{a_{s_i}}^2 + c_{s_{s_i}}^2}{2}\right) \tag{11}$$

where $i = 1, \ldots, N_s$, $\mathbb{U}_{s_i} = \frac{\lambda_{a_{s_i}}}{\mu_{s_i}}$, L_t^s and U_t^s denote the crane pickup and dropoff times.

Electronic copy available at: https://ssrn.com/abstract=2458048

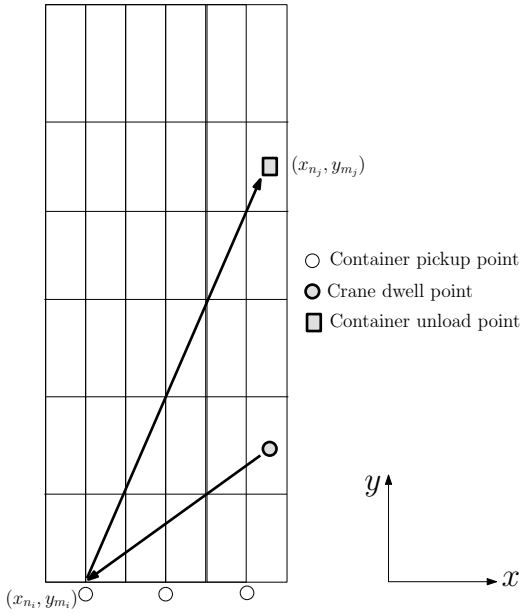

Figure 6 Travel trajectory of a SC during a container unloading process

3.2.4. Integrated Model, Solution Approach, and Performance Measures. Figure 7 describes the integrated queuing model of the container unloading operations from the vessel. On arrival, the containers wait at their respective $GI/G/1$ QC queue $(QC_i, i = 1, \ldots, N_q)$. The SCV of the inter-departure times from the QC queue form the SCV of the inter-arrival times to the vehicle transport synchronization station, J, (see Equation 12, Whitt (1983)). Since there are N_q QCs, the departures from each QC are merged together to form the arrival stream to J. After a vehicle is available, the vehicle and the container are paired and the vehicle joins the IS station. After the service completion, the container joins one of the $GI/G/1$ SC queues $(SC_i, i = 1, \ldots, N_s)$. Therefore, the departure process of the IS station forms the arrival process to the $GI/G/1$ SC queues (Equation 13, Whitt (1983)). Since there are N_s SCs, the departures from the IS station are split into N_s arrival streams, corresponding to each SC queue. After the SC stores the container at a bay location, the container unloading operation is complete.

$$c_{a_t}^2 = \sum_{i=1}^{N_q} \frac{\lambda_{a_{q_i}}}{\lambda_a} c_{d_{q_i}}^2 \tag{12}$$

$$c_{a_{s_i}}^2 = c_{d_t}^2 \left(\frac{1}{N_s}\right) + \left(1 - \frac{1}{N_s}\right) \tag{13}$$

Electronic copy available at: https://ssrn.com/abstract=2458048

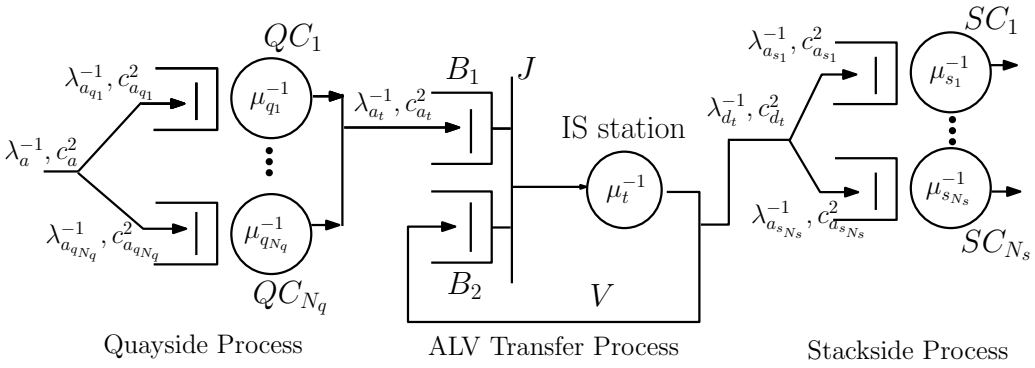

Figure 7 Integrated queuing network model of the container unloading process with ALVs

The $GI/G/1$ queues are solved using Whitt's parametric decomposition techniques (Whitt (1983), Whitt (1994)). One approach to evaluate the vehicle transport $IS-SOQN$ is by using a continuous time Markov chain (CTMC) solution technique. Though the performance measures for the vehicle transport can be obtained from the CTMC, the information on the second moment of the container inter-departure times from the SOQN is also required to link to the downstream SC queue. To estimate the second moment of the container inter-departure times from the SOQN, a detailed analysis of the inter-departure times from the IS station is necessary, which is computationally expensive. To address this issue, we simplify the SOQN by developing an equivalent model of the IS-SOQN.

3.2.5. Equivalent Model of the IS-SOQN and Simplified Model. In this section, we develop a $GI/G/V/\infty$ equivalent model, which can replace the IS-SOQN corresponding to the ALV system. The performance measures and the SCV of the inter-departure times from the ALV network is determined using $GI/G/V$ queue results. We now present the theorem, that states the model equivalencies.

Theorem 3.1 *For an Infinite Server Semi-Open Queuing Network (IS-SOQN) with V vehicles, general inter-arrival times, and general service time distribution, the multi-server $GI/G/V/\infty$ queue provides an exact estimate of the queue length distribution and the throughput at the external queue.*

Electronic copy available at: https://ssrn.com/abstract=2458048

Proof: See Appendix B.

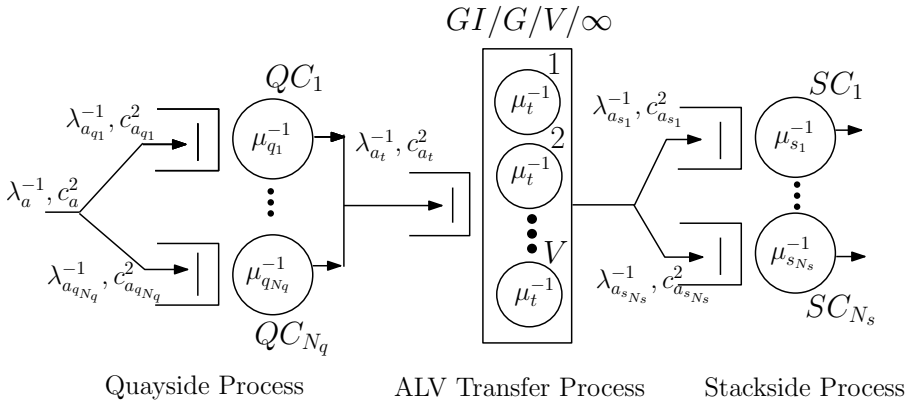

Figure 8 Simplified model of the container unloading process with ALVs

Figure 8 describes the integrated model of the unloading process after replacing the IS-SOQN with a $GI/G/V$ queue. For a $GI/G/V$ queue, the second moment of the departure times and the performance measures can be estimated using parametric decomposition results (Whitt (1993)).

$$c_{d_t}^2 = 1 + (1 - \mathbb{U}_v^2)(c_{a_t}^2 - 1) + \frac{\mathbb{U}_v^2}{\sqrt{V}}(c_{s_t}^2 - 1) \tag{14}$$

$$\mathbb{W}_v = \phi(\mathbb{U}_v, c_{s_t}^2, c_{a_t}^2, V)\left(\frac{u^V \mathbb{U}_v}{V!\lambda_{a_t}(1-\mathbb{U}_v)^2}\right)\left(\frac{c_{s_t}^2 + c_{a_t}^2}{2}\right)p_o \tag{15}$$

where the terms p_o,u, and $\mathbb{U}_v$ are expressed as $\left(\frac{u^V}{V!(1-\mathbb{U}_v)} + \sum_{n=0}^{V-1}\frac{u^n}{n!}\right)^{-1}$, $\frac{\lambda_{a_t}}{\mu_t}$, and $\frac{\lambda_{a_t}}{V\mu_t}$, respectively. The expression for ϕ is present in Whitt (1993).

Little's law can be used to estimate the expected queue lengths (the average number of containers waiting at the quay cranes, ALVs, and stack cranes). The expected throughput time to unload a container ($\mathbb{E}[T_u]$) is provided by Equation 16. The components of the throughput time follows from Equation 1. Note that the two-moment parametric decomposition methods (based on Whitt (1983)) work well under heavy-traffic conditions.

$$\mathbb{E}[T_u] = \mathbb{W}_q(i) + \mu_q^{-1} + \mathbb{W}_v + \mu_t^{-1} + \mathbb{W}_s(i) + \mu_s^{-1} \tag{16}$$

The approach to estimate other performance measures (utilization and expected queue length at all resources) were discussed in Sections 3.2.1-3.2.5.

Electronic copy available at: https://ssrn.com/abstract=2458048

3.2.6. Queuing Network Model for the Loading Operations with ALVs. The integrated queuing network model for the loading operations is illustrated in Figure 9. As described in Figure 3, the process steps in the container loading and unloading operations are identical; however, the order of the process step execution in the container loading and unloading operations are opposite to each other. This aspect is reflected through the routing of containers in the queuing network model. The subnetworks in the queuing network models for the loading and unloading operations are identical. However, the direction of the flow of containers among the subnetworks are opposite to each other. The service time expressions are also identical and the solution approach to evaluate the network is similar. The next section describes the queuing network model for the terminal operations using AGVs.

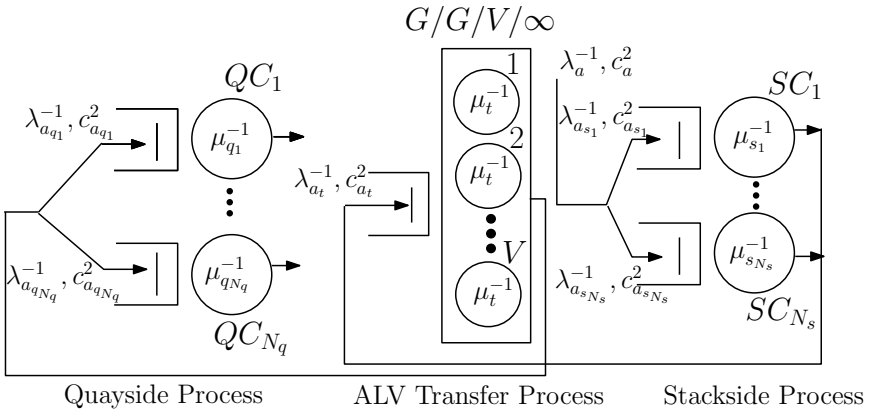

Figure 9 Queuing network model of the container loading process with ALVs

4. Queuing Network Model for Terminal Operations with AGVs

In this section, we develop the model of the unloading operations at a container terminal using AGVs. AGV-based system differs from ALV-based system in terms of the need for *vehicle synchronization* at the quayside and the stackside. In AGV-based system, both QC and the SC drops-off (picks-up) the container on (from) the top of the vehicle. Therefore there is a hard coupling between the vehicle and the QC/SC. We first discuss the protocols that we develop to model the AGV-based terminal operations.

Electronic copy available at: https://ssrn.com/abstract=2458048

1. **Synchronization protocol at the quayside:** For the unloading operation, the QCs begin their operation only when an empty AGV has arrived at the buffer lane to transport the container. Similarly, for the loading operation, the QCs begin their operation only when an AGV loaded with a container has arrived at the quay buffer lane from the stackside.

2. **Synchronization protocol at the stackside:** For the unloading operation, the SCs begin their operation only when an AGV loaded with a container has arrived at the stack buffer lane to store the container. Similarly, for the loading operation, the SCs begin their operation only when an empty AGV has arrived at the stack buffer lane to transport the container to the quayside.

The container unload operation using an AGV is explained now. Due to hard coupling between the AGVs and the QCs, the containers that are waiting to be unloaded need to first wait for an AGV availability (waiting time denoted by W_v). After an AGV is available, it travels to the quayside (travel time denoted by T_{v_1}). Due to the synchronization protocol at the quayside, the AGV waits for the QC to be available and then the QC repositions the container from the vessel to the AGV (the waiting time and repositioning time denoted by W_q and T_q respectively). Then the AGV, loaded with a container, travels to the stackside, waits for a SC availability (synchronization protocol at the stackside). Once a SC is available, the crane travels to the stack buffer lane and picks the container from the AGV. The container is then stored in the stack area. The AGV travel time to the stackside, waiting time for the SC, and the crane travel times are denoted by T_{v_2}, W_s, and T_s respectively. Using these travel and wait time components, the throughput times for the unload and load operations with the AGVs are expressed using Equations 17 and 18 respectively.

$$CT_u = W_v + T_{v_1} + W_q + T_q + T_{v_2} + W_s + T_s \tag{17}$$

$$CT_l = W_v + T_{v_2} + W_s + T_s + T_{v_1} + W_q + T_q \tag{18}$$

4.1. Model Description

The inputs to the queuing network model are the first and second moment of the container inter arrival times (λ_a^{-1}, c_a^2), and the service time information at the resources. The QC and the SC

Electronic copy available at: https://ssrn.com/abstract=2458048

resources are modeled as FCFS stations with general service times. The components of the AGV travel times are modeled as IS stations (VT_1 and VT_2). The AGVs circulate in the network processing container movements.

We now describe the routing of the AGVs and containers in the queuing network model with respect to the unloading operations. The modeling assumptions are unchanged from the ones presented in Section 3.1. Figure 10 describes the queuing network model of the container unloading process with AGVs. The containers that need to be unloaded, wait for an available vehicle at buffer B_1 of the synchronization station J. Idle vehicles wait at buffer B_2. The physical location of the vehicles waiting in buffer B_2 would correspond to the stackside buffer lanes. Once a vehicle and a container is available to be unloaded, then the vehicle queues at the IS station (VT_1). The expected service time at VT_1, $\mu_{t_1}^{-1}$, denotes the expected travel time from its dwell point (point of previous service completion) to the QC buffer lane (Equation 19). After completion of service, the vehicle queues at the QC station ($QC_i, i = 1, \ldots, N_q$) to pick up the container. The expected service time at QC_i, $\mu_{q_i}^{-1}$, denotes the expected movement time of the QC to reach the container in the vessel, container pickup time, movement time to reach the AGV, and container dropoff time. Then, the vehicle queues at the IS station: VT_2. The expected service time at VT_2, $\mu_{t_2}^{-1}$, denotes the expected travel time from the QC buffer lane to the SC buffer lane (Equation 20). After completion of service at VT_2, the vehicle queues at the SC station ($SC_i, i = 1, \ldots, N_s$) to dropoff the container. The expected service time at SC_j, $\hat{\mu}_{s_i}^{-1}$, denotes the expected travel time of the SC from its dwell point to the stack buffer lane and the container pickup time. Once the container is picked up from the AGV, the AGV is now idle and available to transport the containers that are waiting to be unloaded at the quayside.

Note that due to random assignment of containers to a QC and random storage of a container at a stack block, the routing probabilities from station VT_1 to QC_i ($i = 1, \ldots, N_q$) and from VT_2 to SC_i ($i = 1, \ldots, N_s$) are $\frac{1}{N_q}$ and $\frac{1}{N_s}$ respectively. The queuing network in Figure 10 is a semi-open network model because the model possesses the characteristics of both open as well as closed

Electronic copy available at: https://ssrn.com/abstract=2458048

queuing networks. The model is open with respect to the transactions and closed with respect to the vehicles in the network.

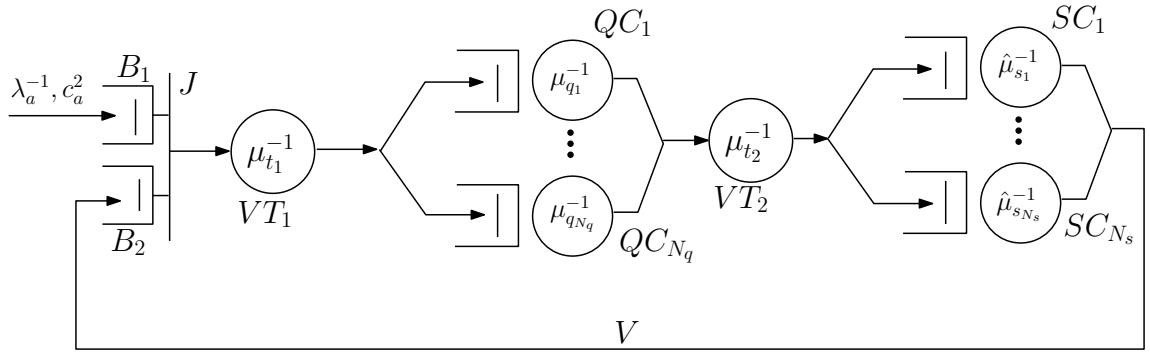

Figure 10 Queuing network model of the container unloading process with AGVs

From the perspective of station SC_i, it is temporarily unavailable whenever it is traveling from the stack buffer lane to store the container at a stack location until the container is repositioned in the stack. To account for the effect of server unavailability, the expected service time at station SC_i is artificially inflated by a suitable factor. Since the proportion of time station SC_i is unavailable for storing a waiting container is $\rho_{s_i}^u = \lambda_{a_{s_i}}(\mathbb{E}[\mathcal{Y}_{bs}] + U_t^s)$, $\hat{\mu}_{s_i}^{-1}$ is determined by Equation 21.

$$\mu_{t_1}^{-1} = \tau_{sq}^{-1} \tag{19}$$

$$\mu_{t_2}^{-1} = \tau_{qs}^{-1} \tag{20}$$

$$\hat{\mu}_{s_i}^{-1} = \frac{\mathbb{E}[\mathcal{Y}_{sb}] + L_t^s}{1 - \rho_{s_i}^u} \tag{21}$$

where $\rho_{s_i}^u = \lambda_{a_{s_i}}(\mathbb{E}[\mathcal{Y}_{bs}] + U_t^s)$. The solution approach and expressions to determine the performance measures are described in Appendix F. After estimating the queue-lengths at all the nodes, the expected throughput time can be estimated using Equation 45. In the next section, we briefly describe the model of the loading operations with AGVs.

4.2. Queuing Network Model for the Loading Operations with AGVs.

Since the flow of containers in the loading process is from stackside to the quaysice, the sequence of the tasks in the loading process are opposite to that of the task sequence in unload operations.

Electronic copy available at: https://ssrn.com/abstract=2458048

Hence, the model developed in Section 4 is also valid for load operations with a variation in the AGV routing. In Figure 10, the location of the idle AGVs (with a point of service completion dwell point) corresponds to quayside buffer lanes. After a container is matched with an AGV, it first queues at IS station, VT_2, receives service at one of the SC stations, queues at IS station VT_1 and travels to the quayside, and finally queues at one of the QC queues to transfer the container in the vessel. The solution approach developed in the previous section can be used to evaluate this model and estimate the performance measures.

5. Model Variation: Analysis of Vehicle Dwell Point

In this section, we investigate the effect of the vehicle dwell point on expected throughput times. We investigate the effect of design parameters using the queuing network model with ALVs. (Refer Appendix E for the analysis of terminal performance with bulk arrivals and insights.) The dwell point of the vehicles indicates the parking location for the vehicles. Vehicles dwell only when containers are not waiting to be unloaded at the instant of the vehicle's job completion instant. The dwell point of the vehicle is an important decision because the vehicle transport times in the yard area are significant. In order to understand the effect of different choices of dwell point on throughput performance, we analyze the vehicle transport system separately. After unloading a container at the stackside, a vehicle can dwell near the stackside, or dwell along the sidepaths, or dwell along the quayside buffer lanes. These three choices are designated as options 1, 2, and 3 in Figure 11.

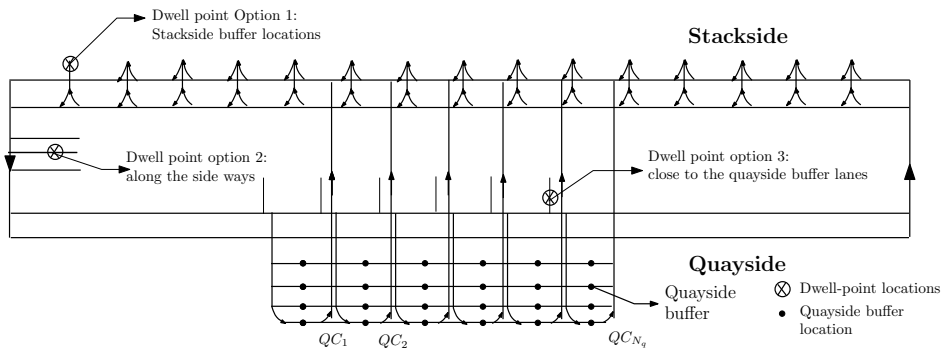

Figure 11 Three options for vehicle dwell point

Electronic copy available at: https://ssrn.com/abstract=2458048

The vehicle transport system is modeled in Section 3.2.2 as an $IS-SOQN$ assuming a point of service completion dwell point. Further, we developed an equivalent model of the vehicle transport system using a $GI/G/V$ queue where the service times denoted the total vehicle travel time from stackside to quayside, container pickup time, vehicle travel time from quayside to stackside, and vehicle dropoff time. This model represented option 1 (point of service completion). However, to model other dwell point choices, the previous modeling description is insufficient. Vehicles travel to a parking location only when there are no containers waiting to be unloaded, i.e, the number of containers waiting in buffer B_1 equals to zero (Figure 7). Since the state-dependent behavior is not captured in the $GI/G/V$ queue, the IS-SOQN model is enriched to capture the state-dependent transitions and a detailed semi-open queuing model is developed.

The SOQN model to model the vehicle dwell point is shown in Figure 12. Let us define a term y, which denotes the difference between the number of containers waiting to be unloaded in buffer B_1 and the number of vehicles idle in buffer B_2 after completing service, $y \in \{-V, \ldots, \infty\}$. There are two classes of vehicles that can be present in buffer B_2: 1) POSC class of vehicles that originated from the stack buffer lanes to unload a container, 2) DWELL class of vehicles that originated from the dwell point location to unload a container. This dwell point depends on the policy option 2 or 3. Depending on the value of the state variable y, vehicles switch classes. For instance, when $y > 0$, a vehicle of class type 'DWELL' switches to class type 'POSC' after completing the unload operation. Table 2 describes the state-dependent switching of vehicle classes.

Table 2 Vehicle class type description for unload operations

Condition on y	Vehicle Class prior to Service	Operation Type	Vehicle Class after Service
$y > 0$	POSC	Unload	POSC
$y > 0$	DWELL	Unload	POSC
$y \leq 0$	POSC	Unload	DWELL
$y \leq 0$	DWELL	Unload	DWELL

Similar to the previous model, the movement of vehicles in the yard area is captured using Infinite Server stations. However, now the routing of the vehicles and the travel times are state-dependent.

Electronic copy available at: https://ssrn.com/abstract=2458048

Hence, the travel time of a vehicle to unload a container is dependent on the originating point of the vehicle (or the vehicle class). If the vehicle class is POSC, then the vehicle visits IS station 1 with an expected service time, μ_t^{-1}, which denotes the time to move from the POSC to the quay crane buffer lane and transport of the container to the stack. After its service completion, the vehicle is routed to station J or station 3 depending on the variable y. If $y \leq 0$, then the vehicle joins IS station 3 to reach the dwell point from the stackside. The expected service time at station 3 is denoted by τ_{sd}^{-1}. If $y > 0$, then the vehicle is routed to station J (buffer B_2) to process another unload operation. Similarly, if the vehicle class is DWELL, it visits IS station 2 with an expected service time, τ_{ds}^{-1}, which denotes the time to move from the dwell point to the quay crane buffer lane and transport of the container to the stack. After its service completion, the DWELL vehicle class routing is identical to POSC vehicle class. The service type expressions for the IS stations, with Option 2 dwell point strategy, are included below.

$$\tau_{sd}^{-1} = \sum_{i=1}^{N_s} \frac{1}{N_s} \left((i-1) \left(\frac{W_s + W_{bs}}{h_v} \right) + \frac{W_l}{2h_v} + \frac{W_s}{2h_v} + \frac{D_{sl} + D_e}{h_v} \right) \tag{22}$$

$$\tau_{dq}^{-1} = \sum_{i=1}^{N_q} \frac{1}{N_q} \left(\frac{W_l}{2h_v} + \frac{L_l}{h_v} + (i-1) \left(\frac{D_{in} + D_{ex}}{h_v} \right) + \frac{W_{bl} + ((N_b - 1)W_{bq})/2}{h_v} + \frac{D_{ex}}{2h_v} \right) \tag{23}$$

Using the expected travel time expressions for τ_{dq}^{-1}, τ_{lu}^{-1}, and τ_{qs}^{-1}, the quantity, τ_{ds}^{-1} is estimated (Equation 24).

$$\tau_{ds}^{-1} = \tau_{dq}^{-1} + \tau_{qs}^{-1} + \tau_{lu}^{-1} \tag{24}$$

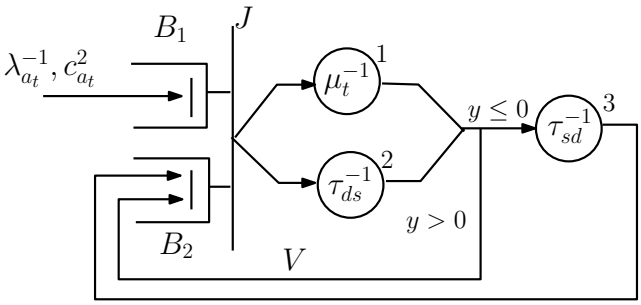

Figure 12 Semi-open queuing network model to analyze vehicle dwell point

Electronic copy available at: https://ssrn.com/abstract=2458048

To evaluate the SOQN with general inter-arrival times (modeled with a Cox-2 arrival process), we develop a CTMC model with a five tuple state space vector (y, p, i, j, k) where y is the difference between the number of transactions waiting at the buffer B_1 and the number of vehicles idle at the buffer B_2, p is the phase of the current unload container arrival, i is the number of vehicles processing transaction from the point of service completion (a stack location) in station 1, j is the number of vehicles processing transaction from the dwell point (DP_1) in station 2, and k is the number of vehicles traveling to the dwell point in station 3. The number of ways that i vehicles are distributed in 3 distinct positions (corresponding to the components: i, j, k of the state vector) is $2\binom{i+3-1}{3-1}$. Note that the factor 2 denotes the arriving transaction can be in Cox phase 1 or 2 ($p \in \{1, 2\}$). Further, when all V vehicles are busy, transactions wait for available vehicles. Since we limit the maximum number of transactions waiting to K, the cardinality of the state space when all vehicles are busy is $\frac{(2K+1)(V+1)(V+2)}{2}$. Hence, the cardinality of the CTMC state space ($|S|$) is given by the expression: $2\sum_{i=0}^{V-1}\binom{i+2}{2} + (2K+1)\binom{V+2}{2}$, where K is the size of the buffer B_1. Upon simplification, $|S| = \frac{(V-1)(V)(V+1)}{3} + \frac{(2K+1)(V+1)(V+2)}{2}$. Table 10 describes the outgoing transitions (to state X_j from state X_i).

The CTMC is solved using the flow balance equations, and the distribution of the vehicles at all nodes are determined. The expected vehicle throughput time is then estimated using the expression $\frac{\mathbb{Q}_{B_1}+\mathbb{Q}_1+\mathbb{Q}_2}{\lambda_a}$, where $\mathbb{Q}_{B_1}$, $\mathbb{Q}_1$, and $\mathbb{Q}_2$ are the mean queue lengths at buffer B_1, node 1 and node 2 respectively.

6. Numerical Experiments and Insights

The results from the analytical model are validated using detailed simulations. The simulation model is build using AutoMod© software v12.2.1 (*www.automod.com*). For each scenario, 15 replications are run with a warmup period of 120 hrs and a run time of 600 hrs ($51,840 - 95,040$ unload containers). The warmup period eliminates any initial bias due to system startup conditions such as the starting location of the vehicles and cranes. Note that the simulation model captures the physical 3D movement of the quay crane, vehicles, and the stack cranes. Further, it also models

Electronic copy available at: https://ssrn.com/abstract=2458048

explicitly the buffer locations at the quay and the stackside. Refer Appendix C for a detailed setup of the simulation model. The design of experiments and the design insights are presented in the following subsections.

6.1. Numerical Experiments

To validate the models in Sections 3 and 4, the QC capacity is set at two levels: 30 cycles/hr and 35 cycles/hr; and the number of vehicles is varied at two levels: 10 and 15 for the ALV system and 15 and 45 for the AGV system. The validation experiments are performed at various utilization levels (70%-90%) of quay crane, vehicles, and stack crane resources. The utilization levels are varied using different arrival rates of the containers. All parameters (layout, vehicle speed and behaviour, stack sizes, number of QCs per vessel) have been determined in cooperation with two large terminal operators (ECT and APMT in Rotterdam) in order to closely resemble terminal operations in practice. A summary of the input parameters for the design of experiments is presented in Table 3. For configuration 1 (with QC capacity 30 cycles/hr and 10 ALVs) and configuration 2 (with QC

Table 3 Design of experiments for model validation (Input) (partly obtained from De Koster et al. (2004))

Configuration/ Area	Quayside	Vehicle Transport	Stackside
Configuration 1	6 Quay cranes	Area of 540m × 90m,	20 stacks, 5 buffer lanes
	capacity: 30 cycles/hr	10 ALVs (30 AGVs)	each stack has 6 rows, 40 bays, 5 tiers
	4 buffer lanes,	velocity: 6 m/s	velocity: 3 m/s
Configuration 2	6 Quay cranes	Area of 540m × 90m,	20 stacks, 5 buffer lanes
	capacity: 35 cycles/hr	15 ALVs (45 AGVs)	each stack has 6 rows, 40 bays, 5 tiers
	4 buffer lanes,	velocity: 6 m/s	velocity: 3 m/s

capacity 35 cycles/hr and 15 ALVs), the arrival rate is varied at 10 levels. Therefore, we simulated 20 scenarios each for ALVs and AGVs, and compared their performance with the analytical model. The average absolute error percentage in expected unload throughput time ($\mathbb{E}[T_u]$), QC utilization ($\mathbb{U}_q$), SC utilization ($\mathbb{U}_s$), vehicle utilization ($\mathbb{U}_v$), and number of containers waiting for the QC ($\mathbb{L}_q$), and SC ($\mathbb{L}_s$) are obtained by the expression ($\left|\frac{A-S}{S}\right| \times 100$), where A and S correspond to the estimate of the measures obtained from analytical and simulation model respectively. The distribution of percentage errors for these measures of interest corresponding to both ALVs and AGVs are summarized in Figures 23 and 24 (see Appendix G). The average percentage errors for

Electronic copy available at: https://ssrn.com/abstract=2458048

the performance measures are included in Table 4. Since the expected number of containers waiting for ALVs, $\mathbb{L}_v$, is low $(0.001 - 0.4)$, we do not report error percentages for the vehicle queue lengths. Similarly, the number of containers waiting for AGVs are < 1. From the error distributions, it can be seen that the ALV model errors are less than 6% for all measures on all scenarios whereas the AGV model errors are less than 11%. In addition, we simulated the ALV system using finite buffers at the quayside and the stackside. For configuration 1, we found that the percentage increase in the ALV throughput time due to finite buffers is only small, 0.8%-1.2%, because there are many buffer locations in total (6/SC and 4/QC).

Table 4 Average percentage errors for the performance measures (Output)

Model Type	**Average % Error**							
	$\mathbb{U}_q$	$\mathbb{L}_q$	$\mathbb{E}[T_q]$	$\mathbb{U}_v$	$\mathbb{E}[T_v]$	$\mathbb{U}_s$	$\mathbb{L}_s$	$\mathbb{E}[T_s]$
ALV	0.2%	2.5%	0.5%	0.6%	0.4%	0.4%	1.7%	3.2%
AGV	0.3%	7.5%	-	1.1%	1.1%	0.4%	7.6%	-

We also validate the analytical model by partnering with an external party, TBA BV (*www.tba.nl*), a company specialized in developing terminal operating system and terminal emulation software. We use data from a container terminal in Virginia, USA, for this validation. The TBA model mimics reality very closely; it includes vehicle congestion, maintains safety distances in the buffer lanes, and considers acceleration and deceleration effects. The comparison results are shown for unloading operations with 6 QCs, 15 and 18 vehicles, and 20 stack blocks. Each stack block contains 40 bays, 10 rows, and 5 tiers. To compare the results, both models were run at very high QC utilizations. We denote the performance measures using a superscript e for the results obtained from the external party, TBA (For details on the model developed by TBA, refer Appendix D). The difference in the vehicle utilization between the TBA model and our model is due to finite buffer capacities and additional space clearance considerations present in the TBA-based model as well as due to slight difference in the QC utilizations between the two models. Based on these validation experiments (Table 5), we conclude that our model can be used to analyze reality with sufficient accuracy.

Electronic copy available at: https://ssrn.com/abstract=2458048

Table 5 Average performance measures (Output) from external party and our model

V	**Comparison on Performance Measures**							
	$\mathbb{U}_v$	$\mathbb{U}_v^e$	$\mathbb{L}_v$	$\mathbb{L}_v^e$	$\mathbb{U}_s$	$\mathbb{U}_s^e$	$\mathbb{L}_s$	$\mathbb{L}_s^e$
15	80%	78%	0.04	0.31	63%	65%	10.8	11.7
18	70%	67%	0.01	0.00	63%	64%	10.9	11.5

6.2. Design Insights

In this section, we present insights with respect to system performance with different vehicle types, vehicle dwell point policy, optimal terminal layout, and effect of batch size on system performance.

6.2.1. Performance Comparison in Coupled vs. Decoupled System. To compare the two system designs, we consider a quay crane with a capacity of 30 cycles/hr (configuration 1, Table 3) and vary the number of vehicles. For each settings with a fixed number of vehicles, the maximum throughput capacity of the system is estimated (see Figure 13). Note that we limit the vehicle utilization to 90%. We find that with the same number of vehicles, the throughput capacity of the ALV-based system is about twice the throughput capacity of AGV-based system. For both systems, we observe that as we increase the number of vehicles, the QCs become a bottleneck. Further, we show the expected throughput time components for ALV (7 vehicles) and AGV-based system (15 vehicles) for an arrival rate of 103 containers/hr (see Figure 14). The waiting time for ALVs is 41 seconds in comparison to 221 seconds in the AGV-based system.

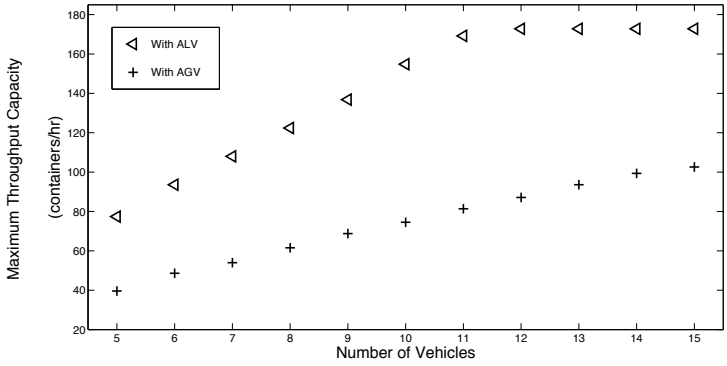

Figure 13 Comparison of throughput capacity of a terminal with ALV (decoupled system) vs. AGV (coupled system)

Electronic copy available at: https://ssrn.com/abstract=2458048

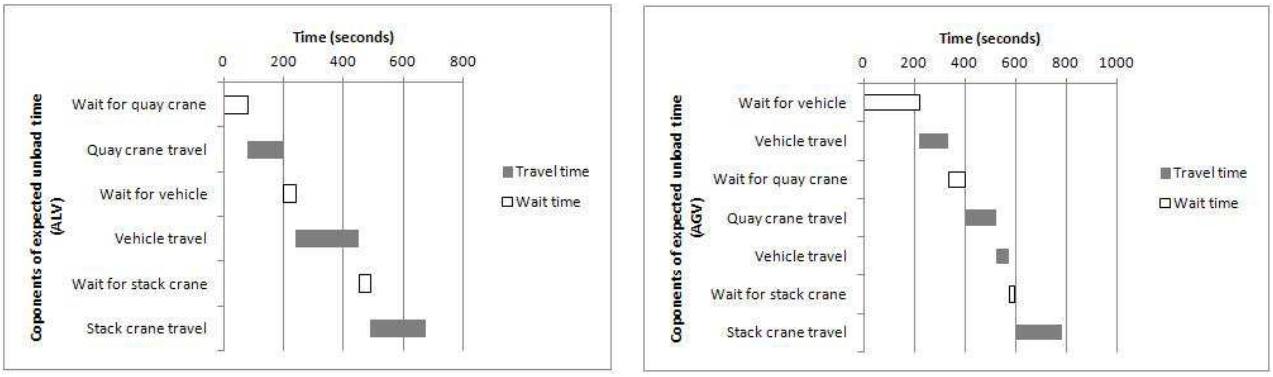

Figure 14 Components of throughput time with ALV (decoupled system) vs. AGV (coupled system) for $\lambda = 103$ containers/hr

Establishing Lower Bounds on the Required Number of ALVs and AGVs: Terminal design engineers can quickly obtain a lower bound of the number of vehicles required for a given throughput. To establish the lower bound for the terminal with ALVs, V_{LB}^{ALV}, we invoke the stability condition on the vehicle network (modeled as a $GI/G/V$ queue). For stability of this multi-server queue, $\frac{\lambda_{a_t}}{V\mu_t} < 1$. By replacing μ_t by $(\tau_{sq}^{-1} + \tau_{lu}^{-1} + \tau_{qs}^{-1})^{-1}$, we get $V > \lambda_{a_t}(\tau_{sq}^{-1} + \tau_{lu}^{-1} + \tau_{qs}^{-1})$. This condition is useful to deploy a minimum number of ALVs for a particular layout configuration.

To establish the lower bound for the number of AGVs, V_{LB}^{AGV}, required for the same arrival rate λ_{a_t}, we further simplify the queuing network with AGVs (shown in Figure 10) by replacing the single server SC and QC queues with Infinite Server (IS) queues. Now the sum of expected service time across the four IS queues is given by $\mu_{t_1}^{-1} + \mu_{q_i}^{-1} + \mu_{t_2}^{-1} + \frac{\mu_{s_i}^{-1}}{2}$. With this approximation, the semi-open queuing network with AGVs can be reduced to a $GI/G/V$ queue (Theorem 3.1). Using the queue stability condition again, we get $V > \lambda_{a_t}(\tau_{sq}^{-1} + \mu_{q_i}^{-1} + \tau_{qs}^{-1} + \frac{\mu_{s_i}^{-1}}{2})$.

$$V_{LB}^{ALV} = \lambda_{a_t}(\tau_{sq}^{-1} + \tau_{lu}^{-1} + \tau_{qs}^{-1}) \tag{25}$$

$$V_{LB}^{AGV} = \lambda_{a_t}(\tau_{sq}^{-1} + \mu_{q_i}^{-1} + \tau_{qs}^{-1} + \frac{\mu_{s_i}^{-1}}{2}) \tag{26}$$

Therefore, we need at least $\left(\frac{V_{LB}^{AGV} - V_{LB}^{ALV}}{V_{LB}^{ALV}}\right) = \left(\frac{\mu_{q_i}^{-1} + \mu_{s_i}^{-1}/2 - \tau_{lu}^{-1}}{(\tau_{sq}^{-1} + \tau_{lu}^{-1} + \tau_{qs}^{-1})}\right) \times 100\%$ more AGVs to achieve the same throughput as obtained by ALVs. For the layout considered in the numerical experiments, we consider a λ_{a_t} of 103 containers/hr. The lower bound of the number of vehicles suggested by

Electronic copy available at: https://ssrn.com/abstract=2458048

the ALV model is 6 whereas the AGV-based model suggests 12. With 90% vehicle utilization, we found that 7 ALVs or 15 AGVs would be needed to achieve the same throughput level.

6.2.2. Dwell Point Policy. To analyze the effect of the dwell point policy, we investigate the vehicle throughput time ($\mathbb{E}[T_v]$) for three dwell point choices (see options 1, 2, and 3 in Figure 11) and compare the performance of dwell-point options 2 and 3 with the point of service completion dwell point (option 1). Let us denote the expected vehicle throughput time for the three options as $\mathbb{E}[T_v^{opt_1}]$, $\mathbb{E}[T_v^{opt_2}]$, and $\mathbb{E}[T_v^{opt_3}]$ respectively.

The percent reduction in vehicle throughput times with variation in dwell point choice 2 (P_{opt_2}) and 3 (P_{opt_3}) are expressed by $P_{opt_2} = \left(\frac{\mathbb{E}[T_v^{opt_1}]-\mathbb{E}[T_v^{opt_2}]}{\mathbb{E}[T_v^{opt_1}]}\right) \times 100\%$ and $P_{opt_3} = \left(\frac{\mathbb{E}[T_v^{opt_1}]-\mathbb{E}[T_v^{opt_3}]}{\mathbb{E}[T_v^{opt_1}]}\right) \times 100\%$ respectively. For both cases, we observe that the percent reduction in expected vehicle throughput times is significant (30%-50% maximum) and the percent difference depends on the choice of dwell point (Figure 15). However, we also observe that the benefit diminishes at high vehicle utilization. As expected, for unload operation, a vehicle dwell point with option 3 is most beneficial.

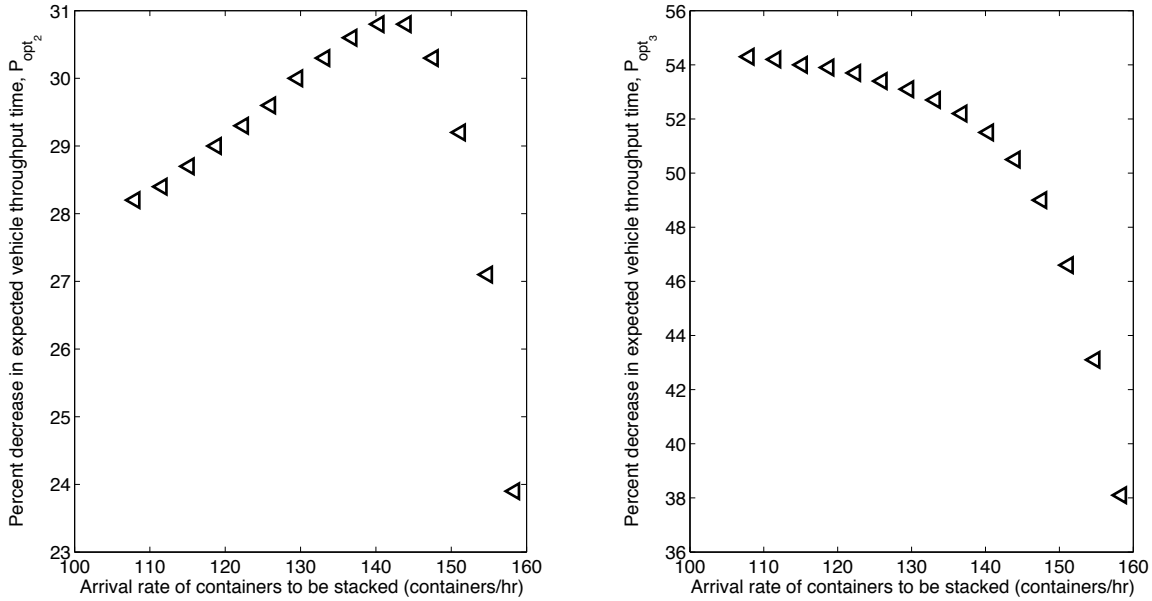

Figure 15 Effect of the dwell point policies on expected vehicle throughput times, Configuration 1 with $V = 10$

Electronic copy available at: https://ssrn.com/abstract=2458048

6.2.3. Terminal Layout Optimization. In this section, we describe the numerical experiments to optimize the layout of the terminal at the stackside with ALVs. To optimize the layout, we assume that the quayside parameters are fixed. For the experiment, we assume $N_q = 6$ and the QCs are positioned symmetrically about the center of the stack blocks. The number of ALVs is fixed at 40. The number of stack blocks, N_s is varied between 20 and 120 with increments of 10. The number of lanes/stack block is varied between 4 and 10 with increments of 2. The number of tiers/stack is either 3 or 5. The number of bays is varied between 20 and 80 such that the total number of container storage locations is about 48,000. With these design constraints, 80 terminal configurations are investigated. For each configuration, we determine the dimensions of the vehicle travel path and estimate the moments of the vehicle travel times. Similarly, at the stackside, we estimate the moments of the crane service times. All configurations are evaluated using the integrated analytical model of ALVs. The computational time on a standard PC for evaluating each configuration is less than a second.

The 80 cases were ranked based on the expected unload throughput time ($\mathbb{E}[T_u]$). Tables 6 and 7 provide insights into two set of design configurations. Information on the bottom five design configurations and the top five designs are included in Tables 6 and 7 respectively. We observe that a large number of stacks with a small number of bays or small number of stacks with large number of bays performs poorly compared with configurations consisting of a small number of stacks with a small number of bays. Since a large number of stacks affects the vehicle travel times, the expected throughput times associated with vehicle travel increase.

Table 6 Poor terminal layout design choices

N_s	N_r	N_b	N_t	$\mathbb{U}_q$	$\mathbb{E}[T_q]$	$\mathbb{U}_v$	$\mathbb{E}[T_v]$	$\mathbb{U}_s$	$\mathbb{E}[T_s]$	$\mathbb{E}[T_u]$
20	8	100	3	67.2%	204.1	22.3%	222.7	87.2%	3727.1	4153.9
30	4	134	3	67.2%	204.1	23.3%	232.6	76.3%	2632.7	3069.3
20	6	80	5	67.2%	204.1	20.0%	199.9	75.2%	1596.5	2000.4
20	10	80	3	67.2%	204.1	24.6%	245.5	71.2%	1322.8	1772.4
110	8	11	5	67.2%	204.1	95.9%	1085.5	3.7%	103.8	1393.4

We explain the interaction among the vehicle transport and the stacking process by investigating their service time relationships. Note that when we increase the number of stack blocks N_s (while

Electronic copy available at: https://ssrn.com/abstract=2458048

Table 7 Good terminal layout design choices

N_s	N_r	N_b	N_t	$\mathbb{U}_q$	$\mathbb{E}[T_q]$	$\mathbb{U}_v$	$\mathbb{E}[T_v]$	$\mathbb{U}_s$	$\mathbb{E}[T_s]$	$\mathbb{E}[T_u]$
30	10	32	5	67.2%	204.1	34.0%	339.9	24.6%	244.0	787.9
40	8	30	5	67.2%	204.1	38.6%	385.9	17.6%	213.5	803.4
30	8	40	5	67.2%	204.1	30.4%	304.1	28.8%	305.1	813.3
40	10	24	5	67.2%	204.1	43.5%	434.6	15.2%	178.8	817.5
40	6	40	5	67.2%	204.1	33.7%	337.1	21.6%	276.7	817.8

keeping the number of storage locations, rows, and tiers constant), the number of bays reduces. As the distance between two random locations in the stack reduces, the expected crane travel time in the stack also reduces. However, the expected vehicle travel distance along the stackside, $(N_s - 1)\frac{W_s + W_{bs}}{h_v}$ and quayside L_l and L_r, also increases. In Figure 16, we vary the number of stack blocks and compare the two cycle times. We fix the number of rows and tiers per stack block to 8 and 5, respectively. Using this comparison, we find 30 stack blocks with 8 rows each gives the lowest total throughput time (third configuration in Table 7). Note that a larger number of stack blocks not only increases throughput time, but also costs, due to the high costs of stacking cranes.

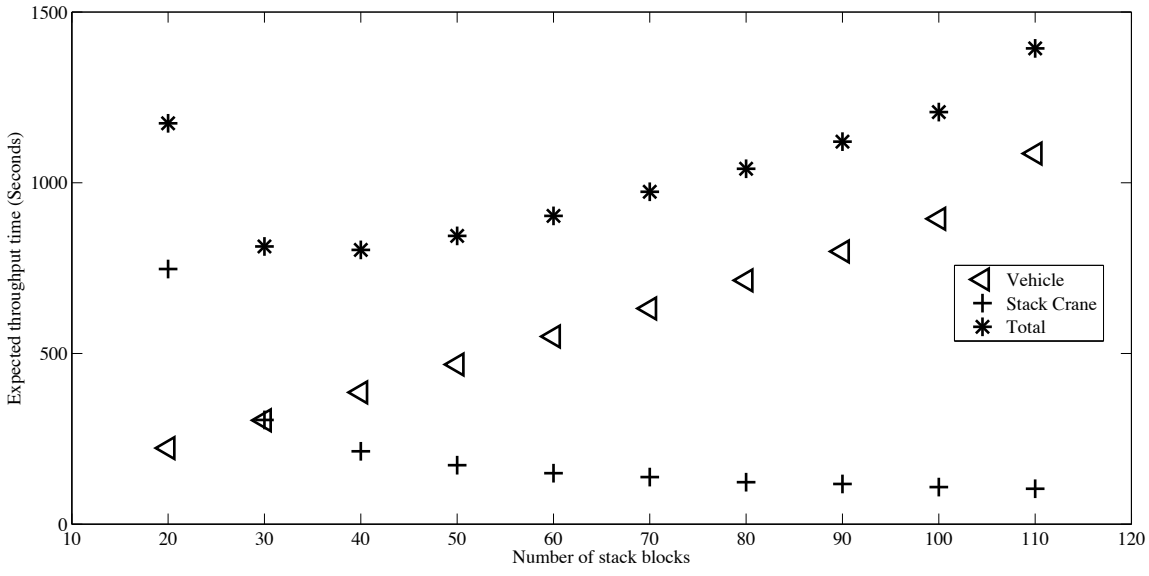

Figure 16 Effect of the number of stack blocks on expected throughput times of vehicles and stack cranes

Electronic copy available at: https://ssrn.com/abstract=2458048

7. Conclusions and Extensions

This research is a first attempt to develop integrated models of the seaside operations by accounting for the dynamic interactions among the quayside, vehicle transport, and stackside processes. Through use of analytical models, we are able to show that an ALV-based system improves the throughput capacity (by 100 %) in comparison to an AGV-based system with the same number of vehicles. Using further model extensions, we are able to show that dwell point of the vehicles (close to the quayside) decreases the expected throughput time by at least 35 %. As expected, the expected unload throughput time increases with increase in container drop size. Numerical experiments also suggest that stack configuration with small number of stacks and small number of bays (20-40 stacks, 20-40 bays) yields better throughput performance than large number of stacks and small number of bays (80 stacks, 20 bays). We believe that the stochastic models of the container handling operations can be used for rapid design conceptualization for container port terminals and improve container handling efficiencies. While we assume uniform assignment of containers at the quay and stackside, the model can be extended to incorporate a skewed distribution of containers assigned to the stack blocks. Further, during the mid-phase of the container handling process, loading and unloading operation are performed simultaneously. Modeling overlapping loading and unloading operations at a container terminal is the subject of future work.

Acknowledgements

We would like to thank the editor and the reviewers for their comments to improve this paper. We also thank SmartPort research centre at Erasmus University (*www.eur.nl/smartport*) for funding this research. We thank the APM Terminal operators at Rotterdam for sharing their insights on terminal design and TBA BV for supporting this paper by providing detailed data and simulation results from a real terminal layout.

Electronic copy available at: https://ssrn.com/abstract=2458048

References

Bae, H., R. Choe, T. Park, K. Ryu, 2011. Comparison of operations of AGVs and ALVs in an automated container terminal. *Journal of Intelligent Manufacturing* **22** 413–426.

Baskett, F., K.M. Chandy, R.R. Muntz, F.G. Palacios, 2005. Open, closed and mixed networks of queues with different classes of customers. *Journal of the ACM* **22** 248–260.

Van den Berg, J.P., 2002. Analytic expressions for the optimal dwell point in an automated storage/retrieval system. *International Journal of Production Economics* **72(1)** 13–25.

Brinkmann, B., 2010. Operations Systems of Container Terminals: A Compendious Overview. *Handbook of Terminal Planning, Eds. Böse, J. W.* Springer Berlin **49** 25–39.

Bolch,G., G. Stefan, H.D. Meer, K.S. Trivedi, 2006. *Queueing Networks And Markov Chains : Modeling and Performance Evaluation with Computer Science Applications*. John Wiley and Sons, Hoboken, NJ.

Canonaco, P., P. Legato, R.M. Mazza, R. Musmanno. 2008. A queuing network model for the management of berth crane operations. *Computers & Operations Research* **35(8)** 2432–2446.

Chandy, K.M., U. Herzog, L. Woo, 1975. Parametric analysis of queuing network models. *IBM J Res* **19(1)** 43–49.

Dekker, R., Voogd, P. and E. Asperen. 2006. Advanced methods for container stacking. *OR Spectrum* **28(4)** 563–586.

De Koster, M.B.M, T. Le-Anh, J.R. van der Meer. 2004. Testing and classifying vehicle dispatching rules in three real-world settings. *Journal of Operations Management* **22(4)** 369–386.

Dragovic, B., N.K. Park, Z. Radmilovic, 2006. Ship-berth link performance evaluation: simulation and analytical approaches. *Maritime Policy & Management* **33(3)** 281–299.

Duinkerken, M., R. Dekker, S. Kurstjens, J. Ottjes, N. Dellaert, 2007. Comparing transportation systems for inter-terminal transport at the Maasvlakte container terminals. *Container Terminals and Cargo Systems, Eds. Kim, K. H., H. Günther, Springer Berlin Heidelberg* 37–61.

Easa, S. M., 1987. Approximate queueing models for analyzing harbor terminal operations. *Transportation Research Part B: Methodological* **21(4)** 269–286.

Edmond, E. D., R. P. Maggs, 1978. How Useful are Queue Models in Port Investment Decisions for Container Berths?. *The Journal of the Operational Research Society* **29(8)** 741–750.

Electronic copy available at: https://ssrn.com/abstract=2458048

Gharehgozli, A.H., G. Laporte, Y. Yu, R. de Koster, 2012. Scheduling Two Yard Cranes Handling Requests with Precedences in a Container Block with Multiple Open Locations. *Working Paper, Rotterdam School of Management, Erasmus University*

www.prweb.com. 2013. Maritime Containerization: A Global Strategic Business Report. Retrieved April 15, 2013, `http://www.prweb.com/releases/containerization/container_shipping/prweb9382752.htm`.

Guan, C., R. Liu, 2009. Container terminal gate appointment system optimization. *Maritime Economics and Logistics* **11(4)** 378–398.

Günther, H.O., K.H. Kim, 2006. Container terminals and terminal operations. *OR Spectrum* **28(4)** 437–445.

Hoshino, S., J. Ota, A. Shinozaki, H. Hashimoto, 2004. Optimal design methodology for an AGV transportation system by using the queuing network theory. *Distributed Autonomous Robotic Systems 6, Eds. Alami, Rachid and Chatila, Raja and Asama, Hajime* Springer Japan 411–420.

Koenigsberg, E., R. C. Lam, 1976. Cyclic queue models of fleet operations. *Operations Research* **24(3)** 516–529.

Kim, K.H., Y. Park, 2004. A crane scheduling method for port container terminals. *European Journal of Operational Research* **156(3)** 752–768.

Kim, K.H., K.Y. Kim , 1999. An optimal routing algorithm for a transfer crane in port container terminals. *Transportation Science* **33(1)** 17–33.

Lazowska, E.D., J. Zahorjan, G.S. Graham, K.C. Sevcik, 1984. *Quantitative System Performance: Computer System Analysis Using Queueing Network Models*. Prentice-Hall, Inc., New Jersey.

Li, W., Y. Wu, M.E.H. Petering, M. Goh, R. de Souza, 2009. Discrete time model and algorithms for container yard crane scheduling. *European Journal of Operational Research* **198(1)** 165–172.

Liang, C., Y. Huang, Y. Yang, 2009. A quay crane dynamic scheduling problem by hybrid evolutionary algorithm for berth allocation planning. *Comput. Ind. Eng.* **56(3)** 1021–1028.

Meisel, F., C. Bierwirth, 2013. A Framework for Integrated Berth Allocation and Crane Operations Planning in Seaport Container Terminals *Transportation Science* **47** 131–147.

Mennis, E., A. Platis, I. Lagoudis, N. Nikitakos, 2008. Improving Port Container Terminal Efficiency with the use of Markov Theory. *Maritime Economics and Logistics* **10(3)** 243–257.

Electronic copy available at: https://ssrn.com/abstract=2458048

Meller, R.D., A. Mungwattana, 2005. AS/RS dwell-point strategy selection at high system utilization: A simulation study to investigate the magnitude of the benefit. *International Journal of Production Research* **43(24)** 5217–5227.

Ng, W.C., 2005. Crane scheduling in container yards with inter-crane interference. *European Journal of Operational Research* **164(1)** 64–78.

Petering, M. E. H., Y. Wu, W. Li, M. Goh, R. de Souza, 2009. Development and simulation analysis of real-time yard crane control systems for seaport container transshipment terminals. *OR Spectrum* **31** 801–835.

Petering, M. E. H., K. G. Murty, 2009. Effect of block length and yard crane deployment systems on overall performance at a seaport container transshipment terminal. *Computers & Operations Research* **36(5)** 1711–1725.

Petering, M. E. H., 2010. Development and simulation analysis of real-time, dual-load yard truck control systems for seaport container transshipment terminals. *OR Spectrum* **32** 633–661.

Petering, M. E. H., 2011. Decision support for yard capacity, eet composition, truck substitutability, and scalability issues at seaport container terminals. *Transportation Research Part E: Logistics and Transportation Review* **47(1)** 85–103.

Saanen, Y. A., R. Dekker, 2007. Intelligent stacking as way out of congested yards? Part 1 *Port Technology International* 87–92.

Son, Y. J. , Wysk, R. A., S. Voß, A.T. Jones. 2003. Simulation-based shop floor control: formal model, model generation and control interface. *IIE Transactions* **35(1)** 29–48.

Steenken, D. , S. Voß, R. Stahlbock. 2004. Container terminal operation and operations research - a classification and literature review. *OR Spectrum* **26(1)** 3–49.

www.freight-int.com. 2010. Design of terminal layout and handling systems. Retrieved December 26, 2011, http://www.freight-int.com/article/design-of-terminal-lay-out-and-handling-systems.html.

Tijms, H.C. 1988. Algorithms and Approximations for Batch-Arrival Queues. Retrieved December 26, 2011, fileadmin/ITCBibDatabase/1988/tijm881.pdf.

Vis, I.F.A., H.J. Carlo, 2010. Sequencing Two Cooperating Automated Stacking Cranes in a Container Terminal. *Transportation Science* **44(2)** 169–182.

Electronic copy available at: https://ssrn.com/abstract=2458048

Vacca, I., Salani, M., M. Bierlaire, 2013. An Exact Algorithm for the Integrated Planning of Berth Allocation and Quay Crane Assignment. *Transportation Science* **47** 148–161.

Vis, I.F.A., R. de Koster, 2003. Transshipment of containers at a container terminal: An overview. *European Journal of Operational Research* **147(1)** 1–16.

Vis, I.F.A., I. Harika. 2004. Comparison of vehicle types at an automated container terminal. *OR Spectrum* **26(1)** 117–143.

Vis, I.F.A., K.J. Roodbergen, 2009. Scheduling of container storage and retrieval. *Operations Research* **57(2)** 456–467.

Whitt, W., 1983. The Queueing Network Analyzer. *Bell System Technical Journal* **62(9)** 2779–2815.

Whitt, W., 1993. Approximations for the GI/G/m queue. *Production and Operations Management* **2(2)** 114–161.

Whitt, W., 1994. Towards better multi-class parametric-decomposition approximations for open queueing networks. *Annals of Operations Research* **48** 221–248.

Wiegmans, B.W., B. Ubbels, P. Rietveld, P. Nijkamp. 2002. Investments in Container Terminals: Public Private Partnerships in Europe *International Journal of Maritime Economics* **4** 1–20.

Appendix A: Vehicle Service Time Expressions

The expressions for the three cases corresponding to the vehicle travel from the quayside to the stackside area are included below.

$$\begin{aligned} \textit{Case 1: } \tau_{qs}(i,j)^{-1} &= \frac{D_{ex}}{2h_v} + \frac{W_{bl} + ((N_b-1)W_{bq})/2}{h_v} + \frac{W_l}{h_v} + \frac{W_{bs}}{2h_v} \\ &+ (N_s - N_{sl}(i) - j)\left(\frac{W_s + W_{bs}}{h_v}\right) + \frac{W_s}{2h_v} + \frac{D_{sl}}{h_v} \\ &\text{for } i = 1,\ldots,N_q, j = 1,\ldots,N_s - N_{sl}(i) \end{aligned} \tag{27}$$

$$\begin{aligned} \textit{Case 2: } \tau_{qs}(i,j)^{-1} &= \frac{D_{ex}}{2h_v} + (j - (N_s - N_{sl}(i)))\left(\frac{D_{in} + D_{ex}}{h_v}\right) + \frac{W_{bl} + ((N_b-1)W_{bq})/2}{h_v} \\ &+ \frac{W_l}{h_v} + \frac{W_s + W_{bs}}{2h_v} + \frac{D_{sl}}{h_v} \\ &\text{for } i = 1,\ldots,N_q - 1, j = N_s - N_{sl}(i) + 1,\ldots,N_s - N_{sl}(i) + N_{sq}(i) - 1 \end{aligned} \tag{28}$$

$$\begin{aligned} \textit{Case 3: } \tau_{qs}(i,j)^{-1} &= \frac{D_{ex}}{2h_v} + (N_{sq} - 1)\left(\frac{D_{in} + D_{ex}}{h_v}\right) + \frac{L_r}{h_v} + \frac{W_{bl} + ((N_b-1)W_{bq})/2}{h_v} + \frac{W_l}{h_v} \\ &+ (N_s - j)\left(\frac{W_s + W_{bs}}{h_v}\right) + \frac{D_e}{h_v} + \frac{W_s}{2h_v} + \frac{D_{sl}}{h_v} \end{aligned} \tag{29}$$

Electronic copy available at: https://ssrn.com/abstract=2458048

$$\text{for } i=1,\ldots,N_q, j=N_s-N_{sl}(i)+N_{sq}(i),\ldots,N_s$$
$$\tau_{qs}^{-1} = \sum_{i=1}^{N_q}\sum_{j=1}^{N_s}\frac{1}{(N_qN_s)}\tau_{qs}(i,j)^{-1} \tag{30}$$

$$\tau_{sq}^{-1} = \sum_{i=1}^{N_q}\sum_{j=1}^{N_s}\frac{1}{(N_q * N_s)}\left(\frac{W_s}{2h_v}+(i-1)\frac{W_s+W_{bs}}{h_v} + \frac{W_l}{h_v}+\frac{L_l}{h_v}+\frac{W_{bl}+((N_b-1)W_{bq})/2}{h_v}+\frac{D_{ex}}{2h_v}+(j-1)\frac{D_{in}+D_{ex}}{h_v}+\frac{D_{sl}+D_e}{h_v}\right) \tag{31}$$

Appendix B: Proof of $GI/G/V$ Queue Equivalency Relationship of an IS-SOQN with Coxian Inputs

Theorem B.1 *For an Infinite Server Semi-Open Queuing Network (IS-SOQN) with V vehicles, general inter-arrival times, and general service time distribution, the multi-server $GI/G/V/\infty$ queue provides an exact estimate of the queue length distribution and the throughput at the external queue.*

Proof: Under the assumption that the distribution of the inter-arrival time and the service time have rational Laplace transforms, we use a Cox-2 distribution to approximate the general inter-arrival times and the service times. Let λ_a^{-1} and c_a^2 denote the first moment and the SCV of the container inter-arrival times. Similarly, let μ_s^{-1} and c_s^2 denote the first moment and the SCV of the container handling times.

Using a Cox-2 distribution, the mean of the two exponential phases of the inter-arrival time are $\lambda_{a_1}^{-1}$ and $\lambda_{a_2}^{-1}$ respectively. Similarly, the mean of the two exponential phases of the service times are $\mu_{s_1}^{-1}$ and $\mu_{s_2}^{-1}$ respectively. The parameters for the Cox-2 inter-arrival times are chosen in the following manner (Bolch et al. (2006)):

$$\lambda_{a_1} = 2\lambda_a \tag{32}$$
$$\lambda_{a_2} = \frac{\lambda_a}{c_a^2} \tag{33}$$
$$p_{a_1} = \frac{1}{2c_a^2} \tag{34}$$

Electronic copy available at: https://ssrn.com/abstract=2458048

Similarly, the parameters for the Cox-2 service times are chosen in the following manner:

$$\mu_{s_1} = 2\mu_s \tag{35}$$

$$\mu_{s_2} = \frac{\mu_s}{c_s^2} \tag{36}$$

$$p_{s_1} = \frac{1}{2c_s^2} \tag{37}$$

We show the equivalency in the two networks by analyzing the CTMC of the IS-SOQN with that of a $C_2/C_2/V$ queue (see Figure 17). Since the two phases of the Coxian distribution are exponential, we do not need to distinguish the phases among servers at the IS station. Let the tuple (n_1, n_2) denote the number of vehicles in the first and the second phase of service at the infinite station, where $(n_1, n_2) \in \{0, 1, \ldots, V\}$. Let the tuple (y, p) denote the difference between number of transactions waiting at buffer B_1 and the number of vehicles waiting for a container at buffer B_2, and phase of the current container in the arrival process, where $y \in \{-V, \ldots, \infty\}$. Therefore, the CTMC for the SOQN can be uniquely defined by the four tuple vector: (y, p, n_1, n_2). Table 8 lists the outgoing transitions for all the possible states in the CTMC (from state X_i to state X_j).

Similarly, the state of a $C_2/C_2/V$ queue can be also uniquely defined by the four-tuple vector (y, p, n_1, n_2). The transitions among the states in this queue is similar to that of the IS-SOQN shown previously. Hence, the two networks are equivalent.

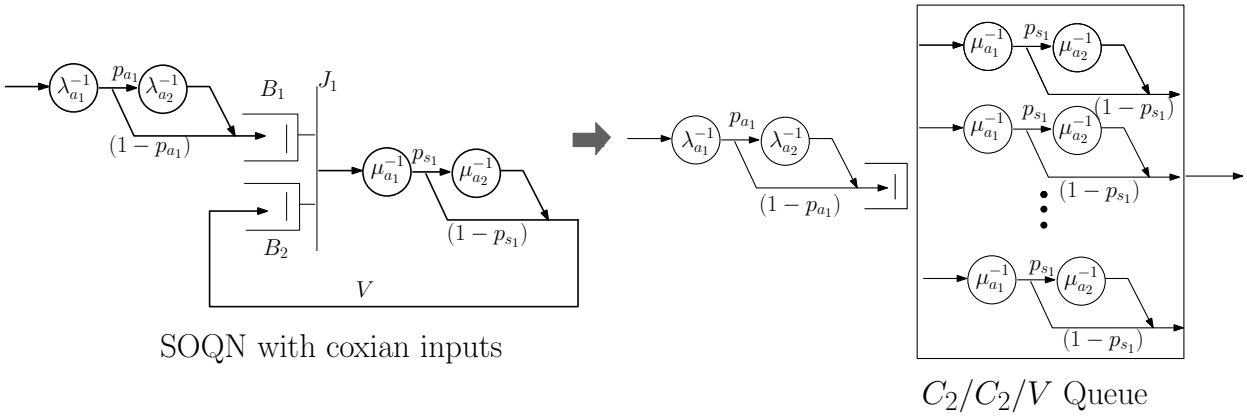

Figure 17 Equivalency of a single station IS-SOQN and a $GI/G/m$ queue

This result can also be shown for generalized phase type distribution inputs.

Electronic copy available at: https://ssrn.com/abstract=2458048

Table 8 Transitions in the CTMC for a single IS-SOQN

Condition	X_i	**Rate**	X_j
$y=-V, p=1, n_1=0, n_2=0$	$(-V,1,0,0)$	$\lambda_{a_1}(1-p_{a_1})$	$(-V+1,1,1,0)$
		$\lambda_{a_1}p_{a_1}$	$(-V,2,0,0)$
$y=-V, p=2, n_1=0, n_2=0$	$(-V,2,0,0)$	λ_{a_2}	$(-V+1,1,1,0)$
$y\geq -V, p=1, n_1>0, n_2=0$	$(y,1,n_1,0)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,n_1+1,0)$
		$\lambda_{a_1}p_{a_1}$	$(y,2,n_1,0)$
		$n_1\mu_{s_1}(1-p_{s_1})$	$(y-1,1,n_1-1,0)$
		$n_1\mu_{s_1}(p_{s_1})$	$(y,1,n_1-1,1)$
$y\geq -V, p=2, n_1>0, n_2=0$	$(y,1,n_1,0)$	λ_{a_2}	$(y+1,1,n_1+1,0)$
		$n_1\mu_{s_1}(1-p_{s_1})$	$(y-1,2,n_1-1,0)$
		$n_1\mu_{s_1}(p_{s_1})$	$(y,2,n_1-1,1)$
$y\geq -V, p=1, n_1=0, n_2>0$	$(y,1,0,n_2)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,1,n_2)$
		$\lambda_{a_1}p_{a_1}$	$(y,2,0,n_2)$
		$n_2\mu_{s_2}$	$(y-1,1,0,n_2-1)$
$y\geq -V, p=2, n_1=0, n_2>0$	$(y,2,0,n_2)$	λ_{a_2}	$(y+1,1,1,n_2)$
		$n_2\mu_{s_2}$	$(y-1,2,0,n_2-1)$
$y\geq -V, p=1, n_1>0, n_2>0$	$(y,1,n_1,n_2)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,n_1+1,n_2)$
		$\lambda_{a_1}p_{a_1}$	$(y,2,n_1,n_2)$
		$n_2\mu_{s_2}$	$(y-1,1,n_1,n_2-1)$
		$n_1\mu_{s_1}(1-p_{s_1})$	$(y-1,1,n_1-1,n_2)$
		$n_1\mu_{s_1}(p_{s_1})$	$(y,1,n_1-1,n_2+1)$
$y\geq -V, p=2, n_1>0, n_2>0$	$(y,2,n_1,n_2)$	λ_{a_2}	$(y+1,1,n_1+1,n_2)$
		$n_2\mu_{s_2}$	$(y-1,2,n_1,n_2-1)$
		$n_1\mu_{s_1}(1-p_{s_1})$	$(y-1,2,n_1-1,n_2)$
		$n_1\mu_{s_1}(p_{s_1})$	$(y,1,n_1-1,n_2+1)$

Appendix C: Description of the In-house Simulation Model

We first describe the components of the simulation model that is developed by the authors and then provide the pseudocode of the simulation model.

C.1. Model components

A simulation model (SM) can be formally defined as the collection of Default Components (DC), Static Information (SI), Dynamic Information (DI), Interaction with Internal/External Modules (IM), Animation (AM), and Statistics Required (SR) i.e., $SM = DC \cup SI \cup DI \cup IEM \cup AM \cup SR$ (see Son et al. (2003)). We describe these components for the simulation models of the container terminal using ALVs that were developed using $AutoMod^{©}$ software.

1. *Default Components (DC)*: To model and run the simulation, we set warm-up period as 5 days, run time as 25 days, and number of replications as 15. These values give us 95% confidence interval values at ± 1% around the mean. The distance was measured in metres and time unit was measured in seconds. In this model, we allow free flow of vehicles without any accumulation in the path. The vehicle handling capacity is one TEU. The velocity of the vehicles and the SC are set at 6m/sec and 3 m/sec respectively.

Electronic copy available at: https://ssrn.com/abstract=2458048

2. *Static Information (SI)*: The static components of the simulation model includes the static stack blocks and the SCs. Each SC in the stackside is modeled with a bridge crane system. The number of bays and length of each bay are 40 TEUs and 12 metres respectively, the velocities of both bridge and the trolley are set at 3m/s and the distance between rails of the bridge are determined based on the number of rows, width of each row and the width of the end rails if the SC. With a width of 2.6 m/row, 6 rows, and 1 m width of the two end rails, we obtain the width of the stack as 17.6 m ($6 \times 2.6 + 2 \times 1$). The distance between each stack block is 10 m. The loading and unloading times of the crane are set at 15 seconds each. Further, we group tiers and bays into Pickup and Dropoff (P&D) points. For instance, a stack with 80 bays and 6 lanes (480 locations per level) are aggregated into 20×2 (40) P&D points by arranging bays and lanes into groups of 4 and 3 respectively. Each container destination location in a stack corresponds to one of the P&D points.

The vehicle transport system, which consists of a set of vehicle travel paths and vehicles, is modeled as a path mover system; the physical paths are drawn based on the layout described in Figure 5. The buffer locations (at quayside and stackside), dwell point (parking) locations, and intermediate travel locations are indicated using control points. The control points are also predefined (static) and cannot be generated dynamically during the simulation run. Each QC is modeled using a bridge crane system. A bridge crane is a set of rails on which a crane moves over pickup and delivery (P&D) areas. The crane moves along the y-axis using rails and the trolley, which is attached to the crane, moves independently along the x-axis. We model the horizontal travel using the crane and trolley movement. However, the vertical travel time is included in the pickup and dropoff times. Each crane operates along four rows and 25 bays. The range of each QC is 20 m.

3. *Dynamic Information (DI)*: The dynamic information for our model includes the container arrival rate and the type of distribution. Also queues are maintained to buffer containers waiting to be handled. Once containers are ready to be unloaded from the vessel, they wait in a queue for QC availability. Likewise, once the containers are unloaded at the SC buffer lanes, the containers wait

Electronic copy available at: https://ssrn.com/abstract=2458048

in a queue for SC availability. At any point in time, an order list is maintained to list the containers waiting for vehicle availability. Once the container arrives, the QC index is assigned to it using a uniform distribution. Then the QC buffer location is chosen with the least number of buffered containers. Using the QC index and the buffer location, the control point of the container origin is defined. Likewise, the SC index of the container is assigned using a uniform distribution. The stack lane buffer location is chosen with the least amount of queued containers. Using the stack crane index and the buffer lane location, the control point for the destination location is formed for the container to be unloaded from the vessel. Row, bay, and tier number of the container storage location are also chosen using discrete uniform distributions.

All resources (QCs, vehicles, and SCs) are supported by order lists (where containers wait and are ordered on their arrival times)

4. *Interaction with Internal/Exteral Modules (IM):* In this simulation model, we do not have the model linked to any external database. However, stochastic interactions are captured between, QC, SC and the vehicle transport system. The containers that are unloaded from the vessel wait in a queue for a vehicle to be available. If a vehicle is not idle at that time instant, the container is listed in an order list. As soon as a vehicle becomes available, it claims the first container from the order list and processes the request. Likewise, the container waits in a queue corresponding to a SC and is listed in an order list, if the crane is unavailable. The crane releases the container from the order list upon availability.

5. *Animation (AM):* The simulation model also allows 3D animation, the speed of animation can be altered, to get a view of the actual system in operation and for debugging. A snapshot from a simulation run is shown in Figure 18.

6. *Statistics Required (SR):* The performance measures of interest are obtained using *AutoStat*© runs. The time of a container arrival and the time at which the container is stored in the stack blocks is stored in the load attributes. Finally, the difference between the completion time and arrival time is recorded in a table. In AutoStat, the responses are defined (such as the average number of containers waiting in a particular QC's order list or a particular SC's orderlist, average

Electronic copy available at: https://ssrn.com/abstract=2458048

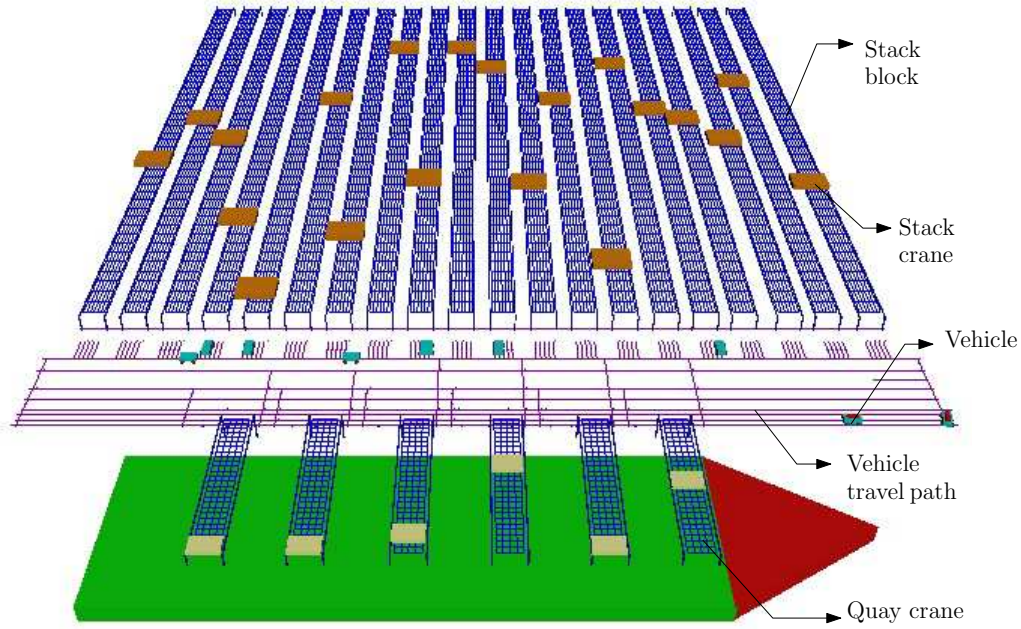

Figure 18 Snapshot of a terminal from the simulation model run in AutoMod

container unload time, vehicle utilization, crane utilization). Using Autostat, the 95% confidence interval for all performance measures are obtained.

The modules present in the simulation model are described in Figure 19. The modules correspond to the three processes of the container terminal operations, quayside, vehicle transport, and stackside, which are modeled as separate systems. The assignment modules assign the resource indices to the containers on arrival. The control modules (at the quay, transport, and the stacksides) are useful for the container dispatch operation. The containers are dispatched only when the resource is available and are controlled by maintaining order lists..

C.2. Pseudocode for unload operations using ALVs

The pseudocode for the container unload operation using ALVs is developed based on the control and unload modules (illustrated in Figure 19). The containers are first dispatched to the `P_ibQuay` process, where the containers are assigned to the QC and SC resource indices. Then the container proceeds to the `P_ship_to_qc` process, where they are repositioned from the vessel to the QC buffer lane. Then the containers are transferred from the quayside to the stackside buffer lane in the `P_qc_to_stack` process. Finally, the containers are stored in the destination stack location

Electronic copy available at: https://ssrn.com/abstract=2458048

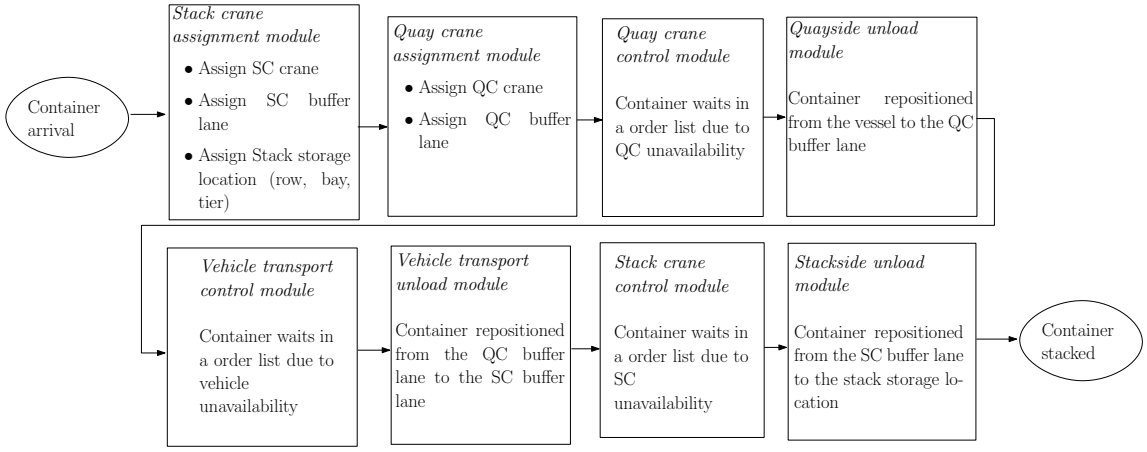

Figure 19 Modules to simulate the container flow process in AutoMod software

in the `P_ibStoStack` process. The load ready and the resource idle procedures are the control procedures for ensuring the flow of containers.

```
//Procedure to assign the QC and SC indices to containers for unloading
begin P_ibQuay arriving procedure
        Assign QC index to the container using discrete uniform distribution
        choose QC buffer location with minimum number of waiting containers
        Assign SC index to the container using discrete uniform distribution
        Assign SC row number using discrete uniform distribution
        Assign SC bay number using discrete uniform distribution
        Assign SC tier number using discrete uniform distribution
        choose SC buffer location with minimum number of waiting containers

        send the container to P_ship_to_qc
end

//Procedure to unload the container from the vessel using the QC
begin P_ship_to_qc arriving procedure
        move the container into an QCQueue at the QC buffer lane
        //Repeat this If condition for all QCs
            if the QC status is not 'Idle' then
                    the container waits to be ordered in OL_QCwait
        move this container into QC origin control point
        travel to the SC destination control point
        move this container into an SCQueue at the SC buffer lane

        send to P_qc_to_stack
end

//Procedure to transport the container from the QC buffer to SC buffer
begin P_qc_to_stack arriving procedure
          if all vehicles are busy then
```


Electronic copy available at: https://ssrn.com/abstract=2458048

```
                    the container waits to be ordered in OL_Vehiclewait
        move into the starting control point at the quayside
        travel to the destination control point at the stackside

        send to P_ibStoStack
end

//Procedure to store the container in the stack
begin P_ibStoStack arriving procedure
        //Repeat this If condition for all SCs
            if the SC status is not ''Idle'' then
                    the container waits to be ordered in OL_SCwait
        move this container into SC origin control point
        travel to the SC destination control point
end

//Procedure to claim the QC resource
//Repeat this procedure for all QCs
begin QC load ready procedure
    the container claims the QC
end

//Procedure to claim the vehicle resource
begin agv load ready procedure
        the container claims one of the idle vehicles
end

//Procedure to claim the SC resource
//Repeat this procedure for all SCs
begin SC load ready procedure
    the container claims the SC
end

//Procedure to release a container waiting for unload
//Repeat this procedure for all QCs
begin QC idle procedure
        if the number of containers waiting in the OL_QCwait > 0 then
                order 1 container from the OL_QCwait to continue
end

//Procedure to release a container waiting for transport
begin vehicle idle procedure
        if the number of containers waiting in the OL_Vehiclewait > 0 then
                order 1 container from OL_Vehiclewait to continue
end

//Procedure to release a container waiting for storage
//Repeat this procedure for all SCs
begin SC idle procedure
```


Electronic copy available at: https://ssrn.com/abstract=2458048

```
        if the number of containers waiting in the OL_SCwait > 0 then
            order 1 container from the OL_SCwait to continue
end
```

Appendix D: Analytical Model Validation from a Real Terminal Simulation

We validated our analytical model with the help of a third-party vendor TBA BV, who specializes in developing Terminal Operating System (TOS), emulation, as well as simulation models for supporting the implementation of container terminals across all continents. In this specific case, we were provided the vehicle and SC utilization along with their expected queue length information for container unload operations from an APM container terminal operating in the east coast of US. The original TBA simulation model that was developed for the container terminal was tuned to accommodate automated lifting vehicles instead of manned lifting vehicles. The simulation model was run for 24 hours The salient features of this model are:

1. The jobs were dispatched by the Terminal Operating Software (TOS), containers were always available to unload (5000 TEUs)

2. The QC service times were obtained from a distribution of service times that account for differences in the container locations in the vessel.

3. The containers were placed in the stack block based on its availability (checking the status of the storage location)

4. The vehicles claimed a buffer lane only when it was free (hence, finite buffer capacities)

5. The SC was not allowed to drop the container adjacent to the container in the SC buffer lanes

6. All resources (cranes, vehicles) have realistic dimensions and behave as in reality. For example, vehicle congestion, acceleration/deceleration are all included in the model.

Appendix E: Analysis of Bulk Container Arrivals

In this Appendix, we describe how the integrated model analysis using ALVs can be extended to bulk container arrivals. When an import vessel arrives and berths, QCs are assigned for the unloading operations. Each crane unloads a batch of containers from the vessel. For simplicity of

Electronic copy available at: https://ssrn.com/abstract=2458048

exposition, we assume that each crane unloads an equal number of containers, $B_q = \lceil \frac{B}{N_q} \rceil$, where B is the number of containers present in the vessel. The QCs are modeled as Markovian Bulk arrival queues, $M^{[X]}/G/1$, where the expected waiting time expression is determined using Equation 38 (Tijms (1988)). This expression can be simplified due to the assumption of constant batch size per QC (Equation 39).

$$\mathbb{W}_q(i) = \left(\frac{\mu_{q_i}^{-1}\rho_{q_i}}{1-\rho_{q_i}}\right)\left(\frac{1+c_{s_{q_i}}^2}{2}\right) + \left(\frac{\mu_{q_i}^{-1}}{2(1-\rho_{q_i})}\right)\left(\frac{\mathbb{E}[\mathcal{X}_b{}^2]}{\mathbb{E}[\mathcal{X}_b]} - 1\right) \tag{38}$$

where $\mathcal{X}_b$ denotes the random variable, batch size per QC. Due to constant batch size assumption, $\mathbb{W}_q(i)$ can be simplified as follows:

$$\mathbb{W}_q(i) = \left(\frac{\mu_{q_i}^{-1}\rho_{q_i}}{1-\rho_{q_i}}\right)\left(\frac{1+c_{s_{q_i}}^2}{2}\right) + \left(\frac{\mu_{q_i}^{-1}\rho_{q_i}}{2(1-\rho_{q_i})}\right)(B_q - 1) \tag{39}$$

Recall that to develop the integrated model of the terminal operations with ALVs, the arrival and departure processes of the quayside, transport, and the stackside need to be linked. Since the departures from the QCs form the arrivals to the vehicle transport, the SCV of the inter-departure times from the QC need to be estimated. We npw describe the approximate approach using embedded Markov chain analysis to analyze the departure process from the bulk arrival queue.

E.1. Departure Process Analysis from Bulk Arrival Queue

Since exact closed-form results that characterize the second moment of inter-departure times from the $M^{[X]}/G/1$ queue are unavailable, one approach would be to study the departure process from the QC station (QC_i) as a Markov renewal process and estimate the mean and SCV of the transaction inter-departure times by analyzing the Markov chain embedded at departure instants from QC_i. However, the number of phases associated with the phase type service time distribution would make the analysis computationally expensive. Hence, we approximate the second moment of inter-departure times (denoted by $\mathbb{E}[\mathcal{X}_d{}^2]$) from an $M^{[X]}/M/1$ by analyzing its embedded Markov chain.

Electronic copy available at: https://ssrn.com/abstract=2458048

First, the transition probability matrix (P_D) is developed and the steady state stationary probability vector (Π_D) is obtained. Using Π_D, the mean and SCV of the inter-departure times from QC_i are obtained. The state of the embedded Markov chain (X_i) has one tuple (j), where $j : j \in \{0, 1, \ldots, \infty\}$ is the number of containers waiting to be unloaded from the vessel by QC_i observed at the departure instant of an unloaded container. For computation of the vector Π_D, we choose a large positive constant K and limit the range of statespace ($S_D = \{0, 1, \ldots, K\}$). The components of P_D can be estimated as follows. When there are one or more containers waiting in the queue to be unloaded, then the probability that the subsequent departure occurs before another batch arrival is given by the expression $\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$ (Note that batch inter-arrival times as well as service times are exponentially distributed with means $\lambda_{a_{q_i}}^{-1}$ and $\mu_{q_i}^{-1}$). Further, the number of batches that arrive before a subsequent departure follows a geometric distribution with success probability, $\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$. Therefore, the probability of n batch arrivals before the subsequent departure is expressed as $\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)^n \left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$. The probability entries for the transitions at the departure instants for $M^{[X]}/M/1$ with B_q=2 is included in Table 9.

Table 9 Markov chain description at the departure instants from $M^{[X]}/M/1$ with B_q=2

State	0	1	2	3	4	...
0	0	$\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	$\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	...
1	$\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	$\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	$\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)^2\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	...
2	0	$\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	$\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	...
3	0	0	$\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	$\left(\frac{\lambda_{a_{q_i}}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	...
4	0	0	0	$\left(\frac{\mu_{q_i}}{\lambda_{a_{q_i}}+\mu_{q_i}}\right)$	0	...
...	...	...	...	...	...	...

Using P_D, the stationary probability vector Π_D of the underlying Markov chain is obtained by solving the system of linear Equations 40 and 41.

$$\Pi_D = \Pi_D P_D \tag{40}$$

$$\sum_{k \in S_D} \Pi_D(X_k) = 1 \tag{41}$$

Electronic copy available at: https://ssrn.com/abstract=2458048

Using the estimates of the steady-state probabilities (Π_D), the estimate of the second moment of the inter-departure times from the bulk-arrival queue is estimated and $c^2_{d_{q_i}}$ is estimated for all QC queues.

$$\mathbb{E}[\chi_d{}^2] = \Pi_D(0)\left(\frac{2}{\lambda^2_{a_{q_i}}} + \frac{2}{\mu^2_{q_i}} + \frac{2}{\mu_{q_i}\lambda_{a_{q_i}}}\right) + \sum_{j=1}^{K}\Pi_D(j)\left(\frac{2}{\mu^2_{q_i}}\right) \tag{42}$$

E.2. Insights with Bulk Container Arrivals.

We consider the ALV-based system (15 vehicles) with bulk container arrivals (50 batch size). The number of batch arrivals vary between 2.88/hr to 3.96/hr. The remaining design parameters are set according to terminal configuration 2. First, we compare the expected throughput time determined from the analytical model with the simulation results. The percentage error range lies between -1.8% to 7% (see Figure 20). The expected number of containers that wait at the quayside vary between 53 to 325 containers.

Further, we compare the throughput time of unload operation by excluding the wait time at the quayside. With a batch size of 50, the percentage difference in the throughput times (with batch) is about 1%-4% more than a single container arrival flow (see Figure 21). Further assuming that the QC is heavily utilized, we can also estimate a lower bound for the expected vessel unloading time by using the expression $\mathbb{E}[T_u] + (B-1)\mu_q^{-1}(i)$, with B as the drop size. As expected, the vessel unload time will increase with the drop size.

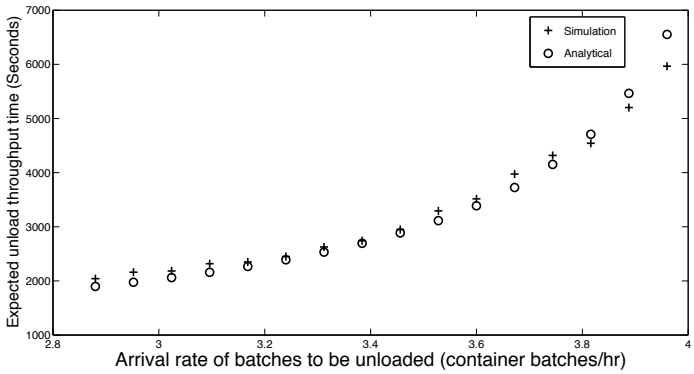

Figure 20 Comparison of expected unload throughput time from analytical and simulation, $V = 15$, $B = 50$

Electronic copy available at: https://ssrn.com/abstract=2458048

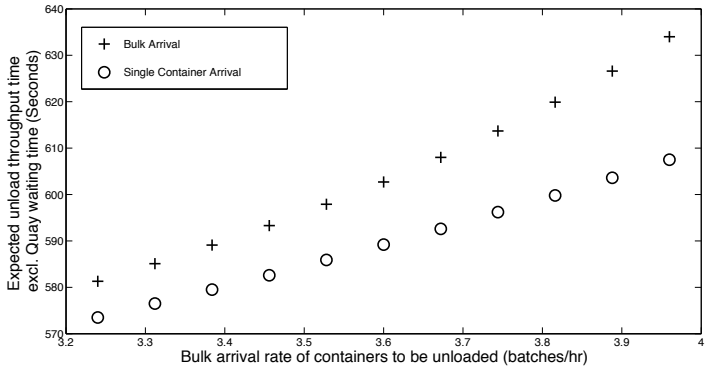

Figure 21 Comparison of expected unload throughput time (with batch vs. single container arrival flow)

Appendix F: Solution Approach for Queuing Network Model with AGVs

Similar to the model with ALVs, the performance measures desired from the queuing model are the average number of vehicles waiting in buffer B_1, QCs, and SCs ($\mathbb{Q}_{B_1}, \mathbb{Q}_{QC_i}$, and $\mathbb{Q}_{SC_i}$), utilization of vehicles, QCs, and SCs ($\mathbb{U}_v, \mathbb{U}_{q_i}$, and $\mathbb{U}_{s_i}$), and expected container throughput times ($\mathbb{E}_{T_v}$). These measures can be obtained from the solution of the queuing network. However, a key challenge is that the queuing network does not have a product-form solution due to presence of a synchronization station (see Baskett et al. (2005)). Therefore, we develop an approximate procedure to evaluate the network and determine performance measures. The four steps of the solution approach are described now.

Network substitution: Let us represent the closed queuing subnetwork (excluding the synchronization station) in the original queuing network model with n vehicles, by a term $\Omega(n)$. We approximate the behavior of subnetwork $\Omega(n)$ with a load dependent queue, L (Chandy et al. (1975)). Note that $\Omega(n)$ is a non BCMP network (due to presence of FCFS stations with general service time distributions, Baskett et al. (2005)). Therefore, this analysis of the reduced SOQN is approximate.

Estimate network parameters: The unknown parameters in the reduced network are the load dependent service rates, $\alpha(n)$, where $n \in \{1, \ldots, V\}$. They are determined by evaluating the throughput of the closed queuing network, $\Omega(n)$ with n vehicles using Approximate Mean Value Analysis (AMVA), i.e., $\alpha(n) = X(n)$.

Electronic copy available at: https://ssrn.com/abstract=2458048

Evaluate the performance of reduced network: The reduced SOQN is evaluated using the continuous time Markov chain (CTMC) with state variable y, which is the difference between the number of containers waiting at buffer B_1 and the number of vehicles idle at buffer, B_2 (see Figure 22). Let $\pi_l(y)$ denote the steady state probabilities of the CTMC.

Determine the performance measures in original network: After estimating the steady state probabilities for the CTMC, the performance measures such as expected queue lengths at the resources are estimated conditional upon the value of the state variable. For instance, the expected queue length at the QC resource, $\mathbb{Q}_{QC_i}$ is determined using Equations 43 and 44. First, the conditional expected queue length, $\mathbb{Q}_{QC_i}|y$, is determined by evaluating the closed queuing network, $\Omega(min(V, y+V))$, where the number of vehicles is expressed as $min(V, y+V)$. Then, the unconditional expected queue length is determined by using the steady state probabilities. The queue lengths at other resources are determined in a similar fashion. The expected throughput times are estimated using Little's law.

$$\mathbb{Q}_{QC_i}|y = \mathbb{Q}_{QC_i}^{\Omega(min(V,y+V))} \tag{43}$$

$$\mathbb{Q}_{QC_i} = \pi_l(y)\mathbb{Q}_{QC_i}|y \tag{44}$$

$$\mathbb{E}[T_u] = \frac{\mathbb{Q}_{B_1}}{\lambda_a} + \frac{\sum_{i=1}^{N_q}\mathbb{Q}_{QC_i} + \sum_{i=1}^{N_s}\mathbb{Q}_{SC_i} + \mathbb{Q}_{VT_1} + \mathbb{Q}_{VT_2}}{\lambda_a} \tag{45}$$

where $\mathbb{Q}_{QC_i}$, $\mathbb{Q}_{SC_i}$, $\mathbb{Q}_{VT_1}$, $\mathbb{Q}_{VT_2}$ denote the expected queue lengths at QC i, SC i, and expected number of vehicles being serviced at nodes VT_1 and VT_2 respectively.

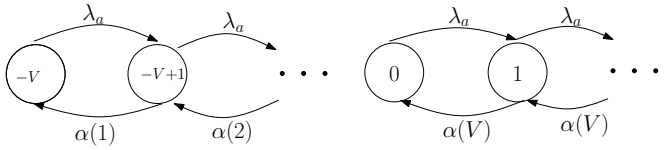

Figure 22 CTMC for evaluating the reduced queuing network with AGVs

Electronic copy available at: https://ssrn.com/abstract=2458048

Appendix G: Summary of Model Errors

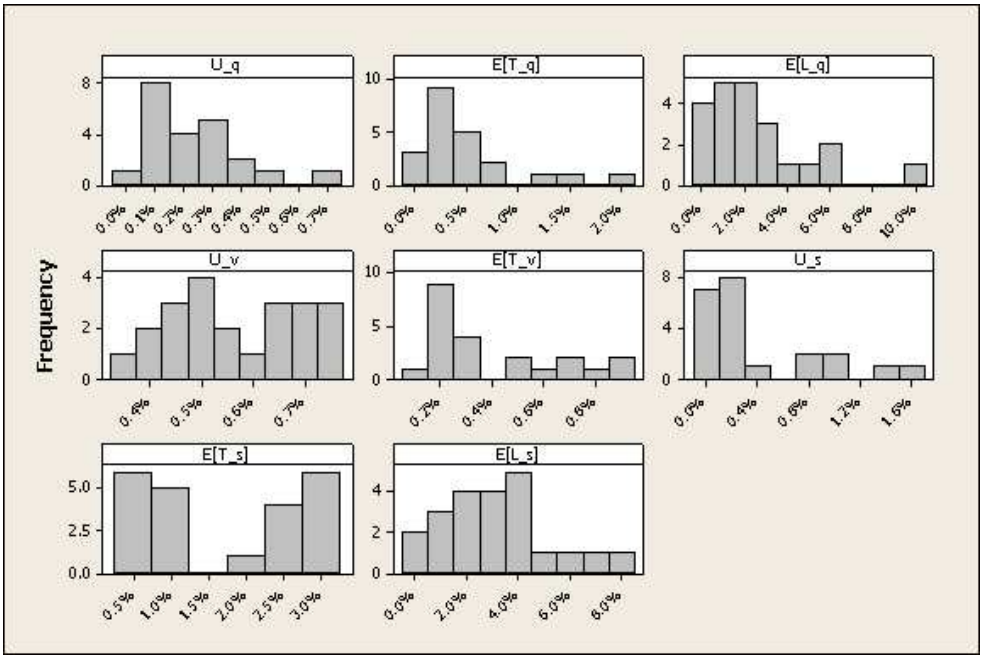

Figure 23 Absolute relative errors for performance measures obtained from ALV model

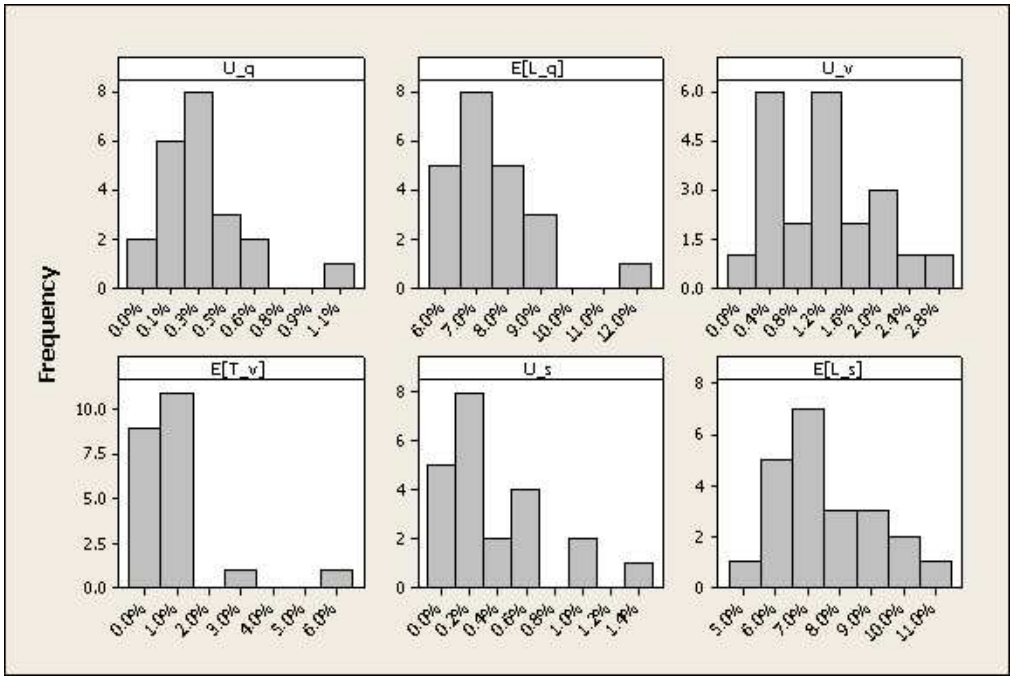

Figure 24 Absolute relative errors for performance measures obtained from AGV model

Electronic copy available at: https://ssrn.com/abstract=2458048

Appendix H: Dwell Point Transitions

Table 10 Transitions in the CTMC for SOQN with state-dependent service times

Condition	X_i	**Rate**	X_j
$y=-V, p=1, i=0, j=0, k=0$	$(-V,1,0,0,0)$	$\lambda_{a_1}(1-p_{a_1})$	$(-V+1,0,1,0)$
		$\lambda_{a_1}(p_{a_1})$	$(-V,2,0,0,0)$
$y=-V, p=2, i=0, j=0, k=0$	$(-V,2,0,0,0)$	λ_{a_2}	$(-V+1,1,1,0,0)$
$-V+1\leq y<0, p=1, i>0, j>0, k>0$	$(y,1,i,j,k)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,i,j+1,k)$
		$\lambda_{a_1}(p_{a_1})$	$(y,2,i,j,k)$
		γ_2	$(y,1,i-1,j,k+1)$
		γ_1	$(y,1,i,j-1,k+1)$
		μ_d	$(y-1,1,i,j,k-1)$
$y=0, p=1, i>0, j>0, k>0$	$(0,1,i,j,k)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,i,j,k)$
		$\lambda_{a_1}(p_{a_1})$	$(y,2,i,j,k)$
		γ_2	$(y,1,i-1,j,k+1)$
		γ_1	$(y,1,i,j-1,k+1)$
		μ_d	$(y-1,1,i,j,k-1)$
$-V+1\leq y<0, p=2, i>0, j>0, k>0$	$(y,2,i,j,k)$	λ_{a_2}	$(y+1,1,i,j+1,k)$
		γ_2	$(y,2,i-1,j,k+1)$
		γ_1	$(y,2,i,j-1,k+1)$
		μ_d	$(y-1,2,i,j,k-1)$
$y=0, p=1, i>0, j>0, k>0$	$(0,2,i,j,k)$	λ_{a_2}	$(y+1,1,i,j,k)$
		γ_2	$(y,2,i-1,j,k+1)$
		γ_1	$(y,2,i,j-1,k+1)$
		μ_d	$(y-1,2,i,j,k-1)$
$y>0, p=1, i>0, j>0, k>0$	$(0,1,i,j,k)$	$\lambda_{a_1}(1-p_{a_1})$	$(y+1,1,i,j,k)$
		$\lambda_{a_1}(p_{a_1})$	$(y,2,i,j,k)$
		γ_2	$(y-1,1,i,j,k)$
		γ_1	$(y-1,1,i+1,j-1,k)$
		μ_d	$(y-1,1,i,j+1,k-1)$
$-V+1\leq y>0, p=2, i>0, j>0, k>0$	$(y,2,i,j,k)$	λ_{a_2}	$(y+1,1,i,j,k)$
		γ_2	$(y-1,1,i,j,k)$
		γ_1	$(y-1,1,i+1,j-1,k)$
		μ_d	$(y-1,1,i,j+1,k-1)$

Electronic copy available at: https://ssrn.com/abstract=2458048

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS–2014–008–LIS
Date of publication	2014–06–23
Version	23–06–2014
Number of pages	59
Persistent URL for paper	http://hdl.handle.net/1765/51536
Email address corresponding author	debjit@iimahd.ernet.in
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus University Rotterdam PO Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31104081182 Fax: +31104089640 Email: info@erim.eur.nl Internet: http://www.erim.eur.nl
Availability	The ERIM Report Series is distributed through the following platforms: RePub, the EUR institutional repository Social Science Research Network (SSRN) Research Papers in Economics (RePEc)
Classifications	The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems: Library of Congress Classification (LCC) Journal of Economic Literature (JEL) ACM Computing Classification System Inspec Classification Scheme (ICS)

Electronic copy available at: https://ssrn.com/abstract=2458048